

DNA Methylation Changes in Clonally Propagated Oil Palm

Norashikin Sarpan^{1¶}, Elizaveta Taranenko^{2,3¶}, Siew-Eng Ooi¹, Eng-Ti Leslie Low¹, Alejandro Espinoza², Tatiana V. Tatarinova^{2,3,4,5*}, Meilina Ong-Abdullah^{1*}

¹ Advanced Biotechnology and Breeding Centre, Malaysian Palm Oil Board, 6 Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia.

² Department of Biology, University of La Verne, La Verne, CA, USA

³ Department of Fundamental Biology and Biotechnology, Siberian Federal University, 660074 Krasnoyarsk, Russia

⁴ Vavilov Institute for General Genetics, Moscow, Russia

⁵ A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

¶ joint first authors

* joint senior/corresponding authors

Emails: Norashikin Sarpan: norashikin.sarpan@mpob.gov.my

Elizaveta Taranenko: taranenko.el@gmail.com

Siew-Eng Ooi: oseng@mpob.gov.my, ORCID: 0000-0002-6621-5456

Eng-Ti Leslie Low: lowengti@mpob.gov.my, ORCID: 0000-0003-3414-8401

Alejandro Espinoza: alejandrosespinoza@laverne.edu

Tatiana V. Tatarinova: ttatarinova@laverne.edu, ORCID: 0000-0003-1787-1112

Meilina Ong-Abdullah: meilina@mpob.gov.my, ORCID: 0000-0003-4825-5021

Abstract

One of the main challenges faced by the oil palm industry is fruit abnormalities, such as the “mantled” phenotype that can lead to reduced yields. This clonal abnormality is an epigenetic phenomenon and has been linked to the hypomethylation of a transposable element within the *EgDEF1* gene. To understand the epigenome changes in clones, methylomes of clonal oil palms were compared to methylomes of seedling-derived oil palms. Whole-genome bisulfite sequencing data from seedlings, normal, and mantled clones were analyzed to determine and compare the context-specific DNA methylomes. In seedlings, coding and regulatory regions are generally hypomethylated while introns and repeats are extensively methylated. Genes with a low number of guanines and cytosines in the third position of codons (GC₃-poor genes) were increasingly methylated towards their 3' region, while GC₃-rich genes remain demethylated, similar to patterns in other eukaryotic species. Predicted promoter regions were generally hypomethylated in seedlings. In clones, CG, CHG, and CHH methylation levels generally decreased in functionally important regions, such as promoters, 5' UTRs, and coding regions. Although random regions were found to be hypomethylated in clonal genomes, hypomethylation of certain hotspot regions may be associated with the clonal mantling phenotype. Our findings, therefore, suggest other hypomethylated CHG sites within the *Karma* of *EgDEF1* and hypomethylated hotspot regions in chromosomes 1, 2, 3 and 5, are associated with mantling.

Keywords

mantling, epigenetics, whole-genome bisulfite sequencing, *Karma*

Declarations

Funding The research was supported by the MPOB

Conflicts of interest/Competing interests None to declare

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication All authors agreed to the publication and approved the final version

Availability of data and material. The central oil palm database is [http://palmxplore.mpob.gov-my/palmXplore/](http://palmxplore.mpob.gov.my/palmXplore/). Whole-genome bisulfite sequence data can be assessed in the NCBI SRA under accession numbers SAMN03569063 - SAMN03569077. The cDNA sequence of the *kDEF1* transcript is available in GenBank under accession number KR347486. RNA-sequencing data can be accessed under GEO Accession No. GSE144079.

Code availability Not applicable

Authors' contributions S-EO and NS performed the experiment, TT, AE, and ET analyzed the data, NS, S-EO, TT, E-TLL, and MO-A interpreted the findings. TT, E-TLL, and MO-A supervised the work. All authors participated in the preparation of the manuscript.

Key message

Several hypomethylated sites within the *Karma* region of *EgDEF1* and hotspot regions in chromosomes 1, 2, 3, and 5 may be associated with mantling.

Abbreviations

WGBS whole-genome bisulfite sequencing
TFBS transcription factor binding site
TE transposable element
mCG methylation at CG context
mCHG methylation at CHG context
mCHH methylation at CHH context

Introduction

Oil palm is the highest oil yielding crop in the world and efforts to increase productivity through the expansion of cultivated land is not sustainable in the long term. A more viable alternative is to increase productivity per unit of the land area using improved planting materials. While the average oil palm plantation yields 4 metric tons per hectare, higher-yielding material from breeding programs can produce over 10 tonnes/hectare (Murphy 2007). Apart from conventional breeding, clones of high yielding palms are an attractive alternative with the potential to increase fresh fruit bunches (FFB) yields by 7 to 34% over seeds (Soh et al. 2011).

Commercial oil palm cloning in Malaysia primarily involves callus initiation and development from leaf explants, followed by somatic embryogenesis and regeneration into plantlets. Two key challenges faced by the oil palm clonal industry are the inefficient cloning process with a low embryogenesis rate and clonal abnormality, in which abnormal (e.g. the mantled phenotype) fruits are produced (Soh et al. 2011) (Fig. 1). In addressing the latter issue, culture media formulations and protocols have undergone massive optimization and evaluation, and their modifications have successfully improved culturing efficiency and mitigated the incidence of mantling (Ong-Abdullah et al. 2015; Hashim et al. 2018). It was found that DNA hypomethylation of a transposable element (TE) within the B-type MADS-box gene *EgDEF1* led to the expression of its transcript isoform, *kDEF1*, in the flowers of mantled clones (Ong-Abdullah et al. 2015). This discovery has enabled oil palm clones to be assayed prior to field planting by gauging their mantling risk. Increasing epigenetic diversity in plants may also improve productivity and create phenotypic plasticity (Latzel et al. 2013; Zhang et al. 2013; Kooke et al. 2015; Fujimoto et al. 2018). Assisted reproductive technology in humans, animals, and plants has been shown to affect the methylome (Nagai et al. 2012; Li et al. 2014; Ghosh et al. 2017; Novakovic et al. 2019).

The tissue culture process leads to methylome changes in crops such as rice and maize (Stroud et al. 2013; Stelpflug et al. 2014; Han et al. 2018). Regenerated plants through tissue culture may provide a source of epigenetic variants for potential breeding or provide planting materials as heterosis potential has been observed from epigenetic recombinant inbred (epiRIL) lines (Latzel et al. 2013; Dapp et al. 2015; Gallusci et al. 2017; Lauss et al. 2018). The use of epigenetic variants as a possible resource for future planting materials may thus require additional tools for marker-assisted breeding such as epi-fingerprinting of planting materials. As epigenetic changes can be reversible, this characteristic also allows for future intervention approaches to curb undesirable impacts on the environment if necessary.

To further understand the epigenome changes in clones for potential epi-breeding efforts in the future, methylomes of clonal oil palms were explored in further detail and compared to methylomes of seed-derived oil palms. Previously, whole-genome bisulfite sequencing (WGBS) data had been generated on seedling, normal, and mantled clonal palms (Ong-Abdullah et al. 2015). We have thus assessed this WGBS data in greater detail to determine the context-specific DNA methylomes of seedling (seed-derived) and clonal oil palms. Methylation differences were also identified in normal clonal palms compared with seedlings to elucidate the methylation changes between seed- and clonal-derived palms. Although hypomethylation of the *Karma* retrotransposon within the *EgDEF1* gene has been linked to mantling (Ong-Abdullah et al. 2015), other differentially methylated regions (DMRs) were also observed in mantled clones compared to both normal clones and seedlings.

Materials and Methods

Plant materials

As reported previously (Ong-Abdullah et al. 2015), leaf tissues of clonal palms exhibiting normal and mantled fruit phenotypes and seed-derived palms of various genetic backgrounds were sampled. In total, leaf tissues from four seedlings, five normal and five mantled clonal palms were used to construct whole-genome bisulfite sequencing (WGBS) libraries. As relevant transcriptome data was not generated previously, leaf tissues from 10-year-old palms of a single clone (five normal and five mantled clonal palms) planted at Malaysian Palm Oil Board research station, Bagan Datuk, Perak, Malaysia, were sampled for transcriptome sequencing.

Oil palm genome and annotation

The *Elaeis guineensis* P5-build of an AVROS *pisifera* palm from Singh et al. (Singh et al. 2013), which contained 40,360 genomic scaffolds (N50 length: 1,045,414 nt; longest length: 22,100,610 nt; and shortest length: 1992 nt) was used. A total of 26,059 high quality predicted gene models (Chan et al. 2017b) obtained using the Seqping algorithm (Chan et al. 2017a) were utilized. The genes are available in the PalmXplore database (Sanusi et al. 2018). The oil palm genomic elements in this study are defined as follows: upstream (2 kb region upstream of the

start codon); 5' UTR (5' untranslated region); CDS (coding sequence containing only exon regions); intron and 3' UTR (3' untranslated region). Due to the insufficient amount of full-length mRNA, prediction of the transcription start sites was not possible for all loci, therefore "upstream" regions in this study include 5' UTR and promoter regions. Gene body constitutes the entire gene sequence, including its introns, with respective predicted TSS and TTS. Intergenic TE refers to TE located in regions between gene bodies and intragenic TE is located within the gene body.

Whole-genome bisulfite sequence data processing

The previously published 100 bp Illumina paired-end sequence reads were used (Ong-Abdullah et al. 2015). All reads within their respective categories (seedling, normal, mantled) were pooled. Quality control checks of the reads were performed by FastQC v0.11.3 software (Babraham Bioinformatics). To remove adapter contaminants and low-quality reads, all reads were trimmed using Trimmomatic v0.36 (Bolger et al. 2014) with the following parameters: *ILLUMINACLIP:TruSeq3-PE,fa:3:13:7:5 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:25 MINLEN:40*. For the mapping of trimmed reads to *Elaeis guineensis* reference genome, Bismark v0.16.3 (Krueger and Andrews 2011) and Bowtie2 v2.2.5 (Langmead and Salzberg 2012) were used with the following Bismark parameters: *--bowtie2 -p 4 -N 1 -L 30 -X 1000 -score_min L,0,-0.6 -sam* and Option *'--directional'* specified (default mode): alignments to complementary strands (CTOT, CTOB) were ignored. Using SAMtools v1.9 (Li et al. 2009), read deduplication was performed by Bismark prior to the identification of methylation state of all cytosines in a CG, CHG and CHH context using the *bismark_methylation_extractor* modules with the following parameters: *-p --no_overlap --bedGraph --buffer_size 20G --scaffolds --cytosine_report --CX --genome_folder*.

Levels of methylation were calculated by Bismark and independently by a custom script. Bismark calculates methylation in a given position as the number of reads reporting methylated C (mC) divided by the number of total reads covering this position. In addition, a Binomial test was used to determine the significance of methylation per position. In this case, methylation of a genomic region is calculated as the ratio of statistically significantly methylated cytosines to all covered cytosines in this region.

RNA isolation, sequencing, and data analysis

Total RNA was isolated using the RNA prep Pure Plant Kit (Tiangen) according to the manufacturer's instructions. Genomic DNA was removed through on-column DNA digestion and the isolated RNA was further purified with RNeasy Plant kit (Qiagen). The quality and quantity of the purified RNA were evaluated using Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries were generated using Kapa stranded mRNA-seq and sequenced on the HiSeq 2500 platform. Raw PE125 reads were checked using FastQC v0.11.3 software (Babraham Bioinformatics) and trimmed with Trim Galore! v0.4.0 (Babraham Bioinformatics). By default, Trim Galore! trims nucleotides with Phred score lower than 20 prior to adapter trimming and discards sequence reads shorter than 20 nt in length. Read mapping to *E. guineensis* reference genome (Singh et al. 2013) was conducted using TopHat v2.0.13 (Kim et al. 2013) with the following parameters: *--b2-very-sensitive --min-intron-length 30 --max-intron-length 50000 -p 10*. Read counting was done using HTSeq v0.6.1p1 (Siretskiy and Spjuth 2014; Wen 2017) and normalized counts were obtained from DESeq2 v3.3.3 (Love et al. 2014).

Statistical analysis

A set of custom scripts was developed to conduct statistical analysis. The *binom.exact* function (R package 'exactci') and difference of proportion test were used for the identification of methylated positions. The R package 'RCircos' was used to draw the circos diagram. Differential methylation in the genomes of clonal (normal and mantled) and seedling samples was calculated using a z-test with a window size of 10,000 nt. Pairwise comparison results were plotted for each chromosome with R packages 'ggplot2' and 'gridExtra', and 'ggpubr'. Confidence intervals were calculated using R. To identify differential methylation of repeats, a z-test was used, and P-values were adjusted using Bonferroni correction.

Transcription factor binding site prediction and analysis

The plant portion of the TRANSFAC database and the MATCH program were used to find potential transcription factor binding sites (TFBSs) (Wingender et al. 1997; Kel et al. 2003). Strin-

gent cut-offs for position weight matrix similarity of 0.95 and the core similarity of 1 were used. A subset of highly informative TFBS for promoter prediction previously generated using the MSU7 rice data (Triska et al. 2017) was applied to oil palm genome analysis. For each TFBS, the distance to the start of translation and position-specific distribution of cytosine methylation was calculated.

Association between methylation and gene expression

To test this association, the *cisExpress* approach (Troukhan et al. 2009; Triska et al. 2013) was used, whereby an average value of gene expression from RNA-Seq data in leaves was computed. Putative TFBS were determined using MATCH (Kel et al. 2003), and a subset of promoter-specific TFBS was identified. The selected region defined as [TSS-1000, TSS+1000] was divided into windows of length W . In each window, the “influence Z-score” of methylation on gene ex-

pression was computed using the following formula: $Z = \frac{(E^M - E^U)}{\sqrt{\frac{V^M}{n^M} + \frac{V^U}{n^U}}}$, where E^M and E^U –

the average expression of genes containing only methylated/unmethylated cytosines in this position; V^M, V^U – variance of gene expression of genes containing methylated/unmethylated cytosines in this position; n^M, n^U number of genes in each category. Since in each window we could only use genes that contain all methylated or all unmethylated C, small window sizes were used, to provide a sufficient number of sequences for analysis.

Results

WGBS had been previously carried out from leaf DNA of seedlings and clones (Ong-Abdullah et al. 2015). The sequenced methylomes were pooled to the designated categories, i.e. seedling, normal, and mantled. A total of 249.54 million, 283.34 million, and 272.9 million raw reads were obtained from seedling, normal, and mantled groups respectively, of which 72-75% were mapped to the *E. guineensis* reference genome.

Methylome landscape of oil palm

In plants, DNA methylation occurs in three contexts: CG, CHG, and CHH (H=C, A or T). The methylomes of the seedling (naturally grown palm) were used to study the global oil palm methylome profile and to establish reliable reference points. In the oil palm genome, 85% of cytosines are covered by at least one read for seedling (88% and 86% for mantled and normal, respectively). Cytosine positions covered by at least 10 reads represent 39%, 43%, and 30% of all genomic cytosines for seedling, mantled, and normal, respectively. The mean coverages for seedling, mantled, and normal was 12X, 13X, and 9.6X. After filtering out positions with zero reads, the mean coverage was 14X, 15X, and 11X for seedling, mantled, and normal. From the Bismark report, the average levels of mCG, mCHG, and mCHH were 74%, 57%, and 5% respectively. This distribution pattern was similar to other plants (Tatarinova et al. 2013).

The oil palm methylome landscape is further shaped at both chromosomal and functional genomic elements. A circos plot representing the chromosome-wide distribution of the different methylation context shows high-density mCG and mCHG regions in all sixteen chromosomes (Fig. 2). In addition, regions with low-density methylation overlapped with gene-rich regions while high-density methylated regions were found at repeat-enriched regions. This indicates that methylation was higher in oil palm repeats than at protein-coding genes. This finding is in accordance with other studies where methylation of TEs, which constitutes mostly repeats, is required for silencing (Cantu et al. 2010). In addition, the oil palm methylome profile across genomic elements revealed that the methylation levels in all three contexts were much higher in the introns, while moderate methylation was identified in coding regions and low methylation levels were observed in the upstream region, 5' and 3' UTRs (Fig. 3). Among the three methylation contexts, mCG levels were the highest across all genomic elements.

DNA methylation was assessed across the gene body and in the 2 kb sequence flanking the genes. A notable decrease of mCG and mCHG levels in the immediate vicinity of the predicted transcription start/termination sites (TSS/TTS) was observed (Fig. 4 and Supplementary Fig. S1). mCG levels were highest in gene bodies. In contrast, methylation in other contexts was relatively low in gene bodies. This observation is consistent with gene body methylation profiles in most plants. The methylation pattern of TEs was also examined, focusing on TEs and their 2 kb flanking regions. Two types of TEs, intragenic (located within genes) and intergenic TEs (located outside of genes) were analyzed. In general, TEs displayed greater methylation levels than the genes (Fig. 4). Intragenic TEs were methylated in all cytosine contexts while intergenic TEs were relatively devoid of mCHH (Fig. 4), suggesting tighter epigenome control (silencing of TE) inside transcribed genes to prevent deleterious effects of expressed TEs.

Oil palm genes have a characteristic right-skewed monocot distribution of GC₃ (fraction of Cs and Gs in the third position of a codon, Supplementary Fig. S2) (Tatarinova et al. 2010; Sablok et al. 2011; Chan et al. 2017b). It has been previously reported that GC₃-rich and -poor genes have different properties; GC₃-rich genes (1) provide more targets for methylation, (2) exhibit more variable expression, (3) more frequently possess upstream TATA boxes, (4) are predominant in certain classes of genes (e.g. stress-responsive genes), (5) have GC₃ content that increases from 5' to 3', and (6) show different methylation gradients (Tatarinova et al. 2010). It has been speculated that opposite patterns of methylation and compositional gradients along coding regions of GC₃-poor (increasing methylation) and GC₃-rich (decreasing methylation) genes formed as a result of several competing processes (Tatarinova et al. 2013).

Similarly, oil palm genes exhibit gradients of GC₃ and DNA methylation patterns differing in the coding regions of GC₃-rich and -poor genes (Fig. 5) (Chan et al. 2017b). The 5'-ends of the coding region of both types of genes (GC₃-rich and -poor) show two effects: (1) elevated levels of GC₃ and (2) reduced levels of DNA methylation. Methylation level in the 5' to 3' direction in

GC₃-poor genes increases, but remains at the same level for approximately 1,000 nt for GC₃-rich genes.

Oil palm regulatory elements are generally demethylated

Core promoters provide blueprints for the assembly of transcriptional machinery through TFBSs. DNA methylation at TFBSs determines the binding affinity of transcription factors at their sequence motifs, thus affecting initiation and regulation of gene transcription. The oil palm putative TFBSs determined with the MATCH program (Kel et al. 2003) showed three types of positional distribution clusters (Fig. 6) that are similar to the profiles of other plants. The three clusters are described as promoter-specific, non-specific, and 5' UTR-specific (Fig. 6a-c), indicating different positional preferences of oil palm transcription factors. mCG dynamics at TFBSs were investigated due to its higher levels compared to mCHG and mCHH. In addition, mCG could serve as an accurate predictor for gene expression (Lioznova et al. 2019). The methylation levels within TFBSs were averaged within 50 nt tiling windows and reported in boxplot to reveal the methylation pattern around the TFBSs (Fig. 6d-e). In promoter-specific TFBSs, methylation level was depleted upstream of TSS while the adjacent 5' UTRs (regions downstream of TSS) were more methylated, most probably to prevent incorrect transcription initiation (Fig. 6d). In 5' UTR-specific TFBSs, depletion of methylation around the 5' UTRs was clearly seen (Fig. 6e). These findings suggest that oil palm TFBSs tend to be demethylated, probably to encourage stable interaction between DNA binding proteins and their target regulatory sequences.

Methylation is assumed to alter gene expression levels, although the relationship between methylation and gene expression is complex (Rishi et al. 2010; Francescatto et al. 2011; Pardo et al. 2013). It has been demonstrated that only a fraction of cytosines exhibits a significant correlation between methylation and expression profiles of a corresponding TSS (Medvedeva et al. 2014). To assess the relationship of gene expression at TFBS regions, a method to quantitatively explore the extent to which DNA methylation at TFBSs can reduce or abolish the expression of adjacent genes was developed. This relationship is described by the Z-score, whereby a positive/negative Z-score value indicates a positive/negative association between gene expression and mCG levels at TFBSs (following the approach by (Troukhan et al. 2009), see Materials and Methods for details). It was found that putative oil palm TFBS regions are enriched with negative Z-score values (Fig. 7). In addition, a cluster of extreme negative Z-score values corresponding to the first 200 nt upstream of the TSS (core promoter) was detected. This indicates the presence of functionally important and epigenetically sensitive regulatory elements in the core promoter, prone to methylation changes and influencing gene transcription. Overall, these findings suggest that TFBS demethylation is likely required for gene expression.

Effects of cloning on methylation

The overall statistics of oil palm WGBS (reported by Bismark) showed that genome-wide methylation levels in mantled clones are lower than in seedlings and normal clones (Table 1). The relative methylation scores for all genes (calculated using a different approach using only statistically significant positions) indicated a significant difference in cytosine methylation between seedlings and clones (Supplementary Fig. S3). Distribution of Z-scores between mantled and normal was generally symmetric while distribution curves between mantled and seedlings or normal and seedlings were right-skewed. This suggests that genome-wide average methylation of all genes is similar for normal and mantled clones; the location of affected regions may explain the phenotypic differences. Hence, the cloning process considerably influenced genome-wide DNA methylation even though there was no phenotypic difference between seedlings and normal clones.

Analysis of the genomic elements showed different levels of methylation among the three groups of oil palm (Supplementary Fig. S4). While differences in CHH methylation between seedlings and normal/mantled clones are minimal, CG and CHG methylation is generally lower in clones compared to seedlings. mCG and mCHG levels were lowest in introns and promoters of normal clones. As shown earlier, the methylation of promoters may influence gene expression, thus reduced promoter methylation in clones may affect gene expression.

TFBSs that were consistently hypo/hypermethylated may reveal regulatory elements that influence expression. Although the overall mCG patterns of TFBSs in promoters are generally similar among seedling, normal, and mantled clonal palms, there are some subtle differences. First of

all, the assessment of differences between promoters of normal and seedlings was required as these differences represented non-functional noise. Differential mCG levels between seedlings and normal palms were found only at four positions, and none of these were within the most functionally important core promoter area since all were located further than 500 nt upstream of the TSS. On the other hand, differential mCG between mantled clones and seedlings was detected at 27 sites and were uniformly distributed from -1000 to +600 of the TSS (i.e., in the promoter and 5' UTR). The corresponding TRANSFAC binding sites are listed in Supplementary Table S1.

The difference of proportions test was conducted to find hypo- and hypermethylated genomic regions in mantled and/or normal clones compared to seedlings. The genome was scanned in windows of 10,000 nt and positions covered by at least 10 reads were used for the analysis. Methylation differences between seedling, mantled and normal clonal palms were absent in most chromosomes, except for parts of chromosomes 1, 2, 3, and 5, which exhibited significant demethylation in mantled clones (Fig. 8).

TE Methylation

Methylation of intergenic and intragenic TEs was compared in seedling, normal and mantled clones (Fig. 9). Average methylation, number of methylated cytosines, and the total number of cytosines were calculated for each type of repeat within genes and between genes. Z-test was applied to identify significantly different TE methylation levels between seedling, normal, and mantled clones. It is evident that methylation levels in mantled clones changed significantly in almost all types of repeats.

As the *Karma* element has been linked with mantling (Ong-Abdullah et al. 2015), methylation was analyzed on the entire *Karma* region as well as its associated transcript isoform, *kDEF1*, and the whole *EgDEF1* gene, whereby the latter spans almost 40 kb, including several introns (Fig. 10). Ong-Abdullah et al. 2015 reported that *Karma* mCHG was markedly reduced in mantled clones while CHH methylation was slightly reduced. Our analysis concurs with this observation as there is a complete absence of *Karma* mCHG in mantled clones (Table S2). *Karma* mCG level was slightly reduced in normal compared to seedlings. In addition, *kDEF1*, which consists of the first five exons of *EgDEF1* and a partial region of the *Karma* element, was also hypomethylated in mantled clones, with decreased mCG and mCHG levels. mCG, mCHG and mCHH levels of *EgDEF1*, which encompasses a much larger region, was only subtly reduced in mantled clones compared to normal clones and seedling.

Discussion

Generally, methylation levels are similar for different tissue types and organs (Hsieh et al. 2009; Zemach et al. 2010; Song et al. 2013), but subtle differences can be detected between, for example, endosperm and vegetative tissues. Several studies have shown that DNA methylation is a dynamic process, as a response to an external stimulus or transition to a different developmental stage (Elhamamsy 2016; Liu et al. 2016). Significant epigenetic changes can happen, for example, during the process of seed development (Pikaard and Mittelsten Scheid 2014; An et al. 2017).

Genome-wide methylation trends in seed-derived oil palms

In the present study relying on WGBS technology, the methylation status of 85% of all cytosines in the oil palm genome was assessed, revealing that 39% of all cytosines were methylated. mCG (74% of all CG sites) made up the largest proportion of the total DNA methylation, followed by mCHG (57%) and mCHH (5%) sites. Similar methylation trends were seen in other flowering plants, with various percentages of methylation frequencies at each cytosine context (Niederhuth et al. 2016; Bartels et al. 2018). Oil palm has overall higher frequencies of mCG and mCHG compared with *Oryza sativa* (58.4% and 31%) and *Arabidopsis thaliana* (30.5% and 10%), but moderately lower than *Zea mays* (86% and 74%). However, the mCHH level in oil palm is comparable to rice (5.5%) and maize (5.4%) but is slightly higher than *Arabidopsis* (3.9%). Widespread variation of DNA methylation levels among plant species has been linked with differences in genomic structure (Niederhuth et al. 2016; Deng et al. 2016). There may be several factors influencing methylation levels, e.g., the amount of gene duplication when an extra gene copy is methylated, the total length of functionally important regions that are protected from methylation (Tatarinova et al. 2013, 2016), and the density of targets for DNA methylation events and repeat content (Piegu et al. 2006; Michael 2014). Interestingly, only the frequency of mCHG was recently found to correlate with genome size (Niederhuth et al. 2016). Different plant species probably choose different strategies to regulate gene expression and use various methylation and demethylation pathways, leading to existing inter-species methylome variation.

In oil palm and other plants, protein-coding genes have low DNA methylation while repeats are extensively methylated, contributing to a high overall level of genome methylation. Among the different genomic elements (excluding extragenic space), intron methylation was higher than any other region in the genome. Intronic methylation could be associated with the increased presence of repeats/TEs in intronic sequences and may be responsible for transposon repression. In maize, TEs within introns were marked by CHG methylation and H3K9me2 (West et al. 2014). However, the effects of intronic methylation on gene expression have been found to depend exclusively on the level of methylation in the first intron position (Anastasiadi et al. 2018).

Similar methylation profiles were seen in switchgrass (Yan et al. 2018) and rice (Tatarinova et al. 2016), but the pattern is different in *Arabidopsis* (Yang et al. 2015) which exhibited higher methylation in exons. One of the explanations for this is in the existence of two classes of genes: GC₃-rich and GC₃-poor genes. Compared to GC₃-poor genes, GC₃-rich genes are shorter, have fewer introns, less gene-body methylation, and more variable gene expression pattern. In grass-

es, the distribution of GC₃ is clearly bimodal, but in the monocot oil palm, GC₃ distribution is right-skewed, exhibiting a long tail of GC₃-rich genes. GC₃-rich and -poor genes have different DNA methylation gradients in gene bodies. In the oil palm, GC₃-poor genes possess increased methylation levels in the 5' to 3' direction, while GC₃-rich genes remain demethylated. This is consistent with the patterns described in rice, human, and even honey bee (Tatarinova et al. 2013).

Although DNA methylation patterns were generally consistent across oil palm samples, certain genomic regions were affected by the cloning process, where epigenetic changes may lead to phenotypic consequences. A pronounced drop in methylation levels was observed in the vicinity of the start and end of transcription is similar to the one described in rice (Triska et al. 2017). In contrast to Wang et al. (Wang et al. 2015), oil palm promoters are less methylated than gene body methylation. However, it is suspected that these disagreements might be due to different definitions of the promoter region.

A negative correlation between gene expression and the level of methylation of certain regulatory elements in oil palm promoters was found. The most negative effect on gene expression comes from methylation of core promoter regions (first 200 nt upstream of TSS). A similar negative correlation between gene expression and promoter methylation was described in soybean (An et al. 2017). Transcription factors have different preferences for binding either to the promoter or the 5' UTR (Triska et al. 2017). The same trend was observed in oil palm. In addition, promoter-specific binding sites in oil palm are less methylated than their surrounding DNA in promoters but not in 5' UTR. The trend for 5' UTR specific TFBSs is reversed, i.e. they are less methylated in the 5' UTR compared to their surrounding sequence.

Cloning leads to genome-wide hypomethylation in oil palm

Tissue culture often causes DNA hypomethylation in callus and regenerated plants (Stroud et al. 2013; Zhang et al. 2014). Regenerated rice plants had ectopic DNA methylation losses enriched at gene promoters, but these were likely a random set of genes that were hypomethylated (Stroud et al. 2013). The rice plants were also significantly enriched with CG and CHG hypomethylated DMRs and overlap in the hypomethylated regions was observed across several lines, suggesting that specific hotspot sites in the genome may be susceptible to methylation losses. The hypomethylation of rice promoters correlates with increased gene expression. Similarly, mCG in oil palm promoters, particularly at TFBS sites, was associated with reduced gene expression. Normal oil palm clones showed a general genome-wide methylation loss at random locations in the genome, compared to seedlings. However, methylation patterns in normal palms were generally similar to seedlings. It appears then that hypomethylation in oil palm normal clones occurs randomly, similar to the general hypomethylation trends observed in regenerated rice (Stroud et al. 2013).

In contrast, mantled clonal palms displayed methylation losses at specific regions of the genome. These methylation differences seemed consistent across several abnormal plants at specific regions of several chromosomes, suggesting that these may be hotspot regions associated with the abnormal phenotype, resembling the methylation hotspots detected in regenerated maize plants (Han et al. 2018). TEs in maize are associated with a substantial number of genes with CHG hypermethylation (West et al. 2014; Niederhuth et al. 2016; Niederhuth and Schmitz 2017). Oil palm TEs are generally hypermethylated but exhibited CHG hypomethylation. Regenerated rice plants also possessed reduced CHG methylation in TEs (Hsu et al. 2018). A high number of differentially methylated TEs was detected in regenerated *indica* rice plants. In Arabidopsis, hypomethylation and TE activation has been observed in long term cell cultures (Tanurdzic et al. 2008). Aberrant demethylation trends observed in mantled clones, therefore, suggest that not only the *Karma* region within *EgDEF1* is affected but parts of several chromosomes also manifested differential methylation. However, it is possible that additional DMRs may not consistently occur in all abnormal palms since the hypomethylation of only the *Karma* region of the *EgDEF1* gene was associated with mantling (Ong-Abdullah et al. 2015). Detailed analysis of *EgDEF1*, including its introns, revealed a complete absence of CHG methylation in the entire *Karma* retrotransposon in mantled clones. In rice, *Karma* was also transcriptionally activated in cultured cells and regenerated plants (Komatsu et al. 2003). Moreover, in regenerated maize plants, differential CHG hypomethylation was detected in genes containing nested TEs, resembling oil palm's *Karma*-like epiallele in *EgDEF1* (Han et al. 2018).

The success rate of re-cloning (“cloning of clones”) is higher than the cloning of seedling (Soh et al. 2011). The genome-wide demethylation detected in clones may contribute to this higher success rate. Additionally, this may indicate that the majority of the genome is sensitive to methylation changes during the standard cloning process of oil palms. The cloning process thus causes epigenetic changes, with methylation losses scattered all over the genome. It has been proposed that DNA methylation is a mechanism to reinforce TE silencing (Lin et al. 2017). TE elements can gain methylation in response to heat stress (Hossain et al. 2017). Several flower development genes are known to be epigenetically regulated. Two epigenetically regulated genes, AGAMOUS and FWA, exhibit methylation changes in prolonged cell suspension cultures of *Arabidopsis thaliana* (Tanurdzic et al. 2008). In melon, the natural presence of a DNA transposon in the CmWIP1 promoter is required for the initiation and maintenance of the spreading of DNA methylation and hypermethylation of the promoter inactivates CmWIP1, allowing carpel development and therefore the formation of female flowers (Martin et al. 2009). Therefore, certain genes associated with floral development may be inclined to methylation changes in response to stress, including stress exposure during cloning. The possible role of *EgDEF1*'s *Karma* in methylation maintenance for proper transcriptional regulation of that region and thus proper floral organ development may be worth further investigation. Mildly mantled clonal palms that can revert to normal phenotype after some time may be suitable samples for such methylation studies.

Acknowledgments

We would like to thank the Director-General of MPOB for permission to publish this study. Our appreciation goes to the Bioinformatics Unit, Breeding and Tissue Culture Unit of MPOB, especially Epigenetics Group and MPOB Bagan Datuk Research Station, Malaysia for their technical support throughout this study. We thank the University of La Verne for the support of this international collaboration.

References

- Anastasiadi D, Esteve-Codina A, Piferrer F (2018) Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* 11:37. <https://doi.org/10.1186/s13072-018-0205-1>
- An Y-QC, Goettel W, Han Q, et al (2017) Dynamic changes of genome-wide DNA methylation during soybean seed development. *Sci Rep* 7:12263. <https://doi.org/10.1038/s41598-017-12510-4>
- Bartels A, Han Q, Nair P, et al (2018) Dynamic DNA methylation in plant growth and development. *Int J Mol Sci* 19:2144. <https://doi.org/10.3390/ijms19072144>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Cantu D, Vanzetti LS, Sumner A, et al (2010) Small RNAs, DNA methylation and transposable elements in wheat. *BMC Genomics* 11:408. <https://doi.org/10.1186/1471-2164-11-408>
- Chan K-L, Rosli R, Tatarinova TV, et al (2017a) Seqping: gene prediction pipeline for plant genomes using self-training gene models and transcriptomic data. *BMC Bioinformatics* 18:1426. <https://doi.org/10.1186/s12859-016-1426-6>
- Chan K-L, Tatarinova TV, Rosli R, et al (2017b) Evidence-based gene models for structural and functional annotations of the oil palm genome. *Biol Direct* 12:21. <https://doi.org/10.1186/s13062-017-0191-4>
- Dapp M, Reinders J, Bédiée A, et al (2015) Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. *Nat Plants* 1:15092. <https://doi.org/10.1038/nplants.2015.92>
- Deng X, Song X, Wei L, et al (2016) Epigenetic regulation and epigenomic landscape in rice. *Natl Sci Rev* 3:309–327. <https://doi.org/10.1093/nsr/nww042>
- Elhamamsy AR (2016) DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation. *Cell Biochemistry and Function* 34:289–298. <https://doi.org/10.1002/cbf.3183>
- Francescato M, Pardo L, Rizzu P, et al (2011) Profiling transcription initiation in human aged brain using deep-CAGE. *BMC Bioinformatics* 12:A8. <https://doi.org/10.1186/1471-2105-12-S11-A8>
- Fujimoto R, Uezono K, Ishikura S, et al (2018) Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breed Sci* 68:145–158. <https://doi.org/10.1270/jsbbs.17155>
- Gallusci P, Dai Z, Génard M, et al (2017) Epigenetics for plant improvement: current knowledge and modeling avenues. *Trends Plant Sci* 22:610–623. <https://doi.org/10.1016/j.tplants.2017.04.009>
- Ghosh J, Coutifaris C, Sapienza C, Mainigi M (2017) Global DNA methylation levels are altered by modifiable clinical manipulations in assisted reproductive technologies. *Clin Epigenetics* 9:14. <https://doi.org/10.1186/s13148-017-0318-6>
- Han Z, Crisp PA, Stelpflug S, et al (2018) Heritable epigenomic changes to the maize methylome resulting from tissue culture. *Genetics* 209:983–995. <https://doi.org/10.1534/genetics.118.300987>
- Hashim AT, Ishak Z, Rosli SK, et al (2018) Oil palm (*Elaeis guineensis* Jacq.) somatic embryogenesis. In: Jain SM, Gupta P (eds) Step wise protocols for somatic embryogenesis of important woody plants. Springer International Publishing, pp 209–229
- Hossain MS, Kawakatsu T, Kim KD, et al (2017) Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs. *New Phytol* 214:808–819. <https://doi.org/10.1111/nph.14421>

- Hsieh T-F, Ibarra CA, Silva P, et al (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* 324:1451–1454. <https://doi.org/10.1126/science.1172417>
- Hsu F-M, Gohain M, Allishe A, et al (2018) Dynamics of the methylome and transcriptome during the regeneration of rice. *Epigenomes* 2:14. <https://doi.org/10.3390/epigenomes2030014>
- Kel AE, Gössling E, Reuter I, et al (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 31:3576–3579. <https://doi.org/10.1093/nar/gkg585>
- Kim D, Pertea G, Trapnell C, et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>
- Komatsu M, Shimamoto K, Kyojuka J (2003) Two-Step Regulation and Continuous Retrotransposition of the Rice LINE-Type Retrotransposon *Karma*. *The Plant Cell* 15:1934–1944. <https://doi.org/10.1105/tpc.011809>
- Kooke R, Johannes F, Wardenaar R, et al (2015) Epigenetic basis of morphological variation and phenotypic plasticity in Arabidopsis thaliana. *Plant Cell* 27:337–348. <https://doi.org/10.1105/tpc.114.133025>
- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Latzel V, Allan E, Bortolini Silveira A, et al (2013) Epigenetic diversity increases the productivity and stability of plant populations. *Nat Commun* 4:2875. <https://doi.org/10.1038/ncomms3875>
- Lauss K, Wardenaar R, Oka R, et al (2018) Parental DNA Methylation States Are Associated with Heterosis in Epigenetic Hybrids. *Plant Physiol* 176:1627–1645. <https://doi.org/10.1104/pp.17.01054>
- Li G, Jia Q, Zhao J, et al (2014) Dysregulation of genome-wide gene expression and DNA methylation in abnormal cloned piglets. *BMC Genomics* 15:811. <https://doi.org/10.1186/1471-2164-15-811>
- Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAM-tools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lin J-Y, Le BH, Chen M, et al (2017) Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. *PNAS* 114:E9730–E9739. <https://doi.org/10.1073/pnas.1716758114>
- Lioznova AV, Khamis AM, Artemov AV, et al (2019) CpG traffic lights are markers of regulatory regions in human genome. *BMC Genomics* 20.: <https://doi.org/10.1186/s12864-018-5387-1>
- Liu H, Li S, Wang X, et al (2016) DNA methylation dynamics: identification and functional annotation. *Brief Funct Genomics* 15:470–484. <https://doi.org/10.1093/bfgp/elw029>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15:
- Martin A, Troadec C, Boualem A, et al (2009) A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461:1135–1138. <https://doi.org/10.1038/nature08498>
- Medvedeva YA, Khamis AM, Kulakovskiy IV, et al (2014) Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics* 15:119. <https://doi.org/10.1186/1471-2164-15-119>
- Michael TP (2014) Plant genome size variation: bloating and purging DNA. *Brief Funct Ge-*

nomics 13:308–317. <https://doi.org/10.1093/bfpg/elu005>

- Murphy DJ (2007) Future prospects for oil palm in the 21st century: Biological and related challenges. *Eur J Lipid Sci Technol* 109:296–306. <https://doi.org/10.1002/ejlt.200600229>
- Nagai M, Meguro-Horike M, Horike S-I (2012) Epigenetic defects related reproductive technologies: Large offspring syndrome (LOS). In: Tatarinova T (ed) DNA methylation - from genomics to technology. InTech Open, p 167
- Niederhuth CE, Bewick AJ, Ji L, et al (2016) Widespread natural variation of DNA methylation within angiosperms. *Genome Biology* 17:194. <https://doi.org/10.1186/s13059-016-1059-0>
- Niederhuth CE, Schmitz RJ (2017) Putting DNA methylation in context: from genomes to gene expression in plants. *Biochim Biophys Acta Gene Regul Mech* 1860:149–156. <https://doi.org/10.1016/j.bbagr.2016.08.009>
- Novakovic B, Lewis S, Halliday J, et al (2019) Assisted reproductive technologies are associated with limited epigenetic variation at birth that largely resolves by adulthood. *Nat Commun* 10:3922. <https://doi.org/10.1038/s41467-019-11929-9>
- Ong-Abdullah M, Ordway JM, Jiang N, et al (2015) Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* 525:533–537. <https://doi.org/10.1038/nature15365>
- Pardo LM, Rizzu P, Francescato M, et al (2013) Regional differences in gene expression and promoter usage in aged human brains. *Neurobiol Aging* 34:1825–1836. <https://doi.org/10.1016/j.neurobiolaging.2013.01.005>
- Piegu B, Guyot R, Picault N, et al (2006) Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16:1262–1269. <https://doi.org/10.1101/gr.5290206>
- Pikaard CS, Mittelsten Scheid O (2014) Epigenetic regulation in plants. *Cold Spring Harb Perspect Biol* 6:a019315. <https://doi.org/10.1101/cshperspect.a019315>
- Rishi V, Bhattacharya P, Chatterjee R, et al (2010) CpG methylation of half-CRE sequences creates C/EBP binding sites that activate some tissue-specific genes. *Proceedings of the National Academy of Sciences* 107:20311–20316. <https://doi.org/10.1073/pnas.1008688107>
- Sablok G, Nayak KC, Vazquez F, Tatarinova TV (2011) Synonymous codon usage, GC₃, and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Mol Biotechnol* 49:116–128. <https://doi.org/10.1007/s12033-011-9383-9>
- Sanusi NSNM, Nik Shazana Nik, Rosli R, et al (2018) PalmXplore: oil palm gene database. *Database* 2018:bay095. <https://doi.org/10.1093/database/bay095>
- Singh R, Ong-Abdullah M, Low E-TL, et al (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500:335–339. <https://doi.org/10.1038/nature12309>
- Siretskiy A, Spjuth O (2014) HTSeq-Hadoop: Extending HTSeq for massively parallel sequencing data analysis using Hadoop. In: 2014 IEEE 10th International Conference on e-Science. pp 317–323
- Soh AC, Wong G, Tan CC, et al (2011) Commercial-scale propagation and planting of elite oil palm clones: research and development towards realization. *J Oil Palm Res* 23:935–952
- Song Q-X, Lu X, Li Q-T, et al (2013) Genome-wide analysis of DNA methylation in soybean. *Mol Plant* 6:1961–1974. <https://doi.org/10.1093/mp/sst123>
- Stelpflug SC, Eichten SR, Hermanson PJ, et al (2014) Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. *Genetics* 198:209–218. <https://doi.org/10.1534/genetics.114.165480>

- Stroud H, Ding B, Simon SA, et al (2013) Plants regenerated from tissue culture contain stable epigenome changes in rice. *Elife* 2:e00354. <https://doi.org/10.7554/eLife.00354>
- Tanurdzic M, Vaughn MW, Jiang H, et al (2008) Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol* 6:e302. <https://doi.org/10.1371/journal.pbio.0060302>
- Tatarinova T, Elhaik E, Pellegrini M (2013) Cross-species analysis of genic GC₃ content and DNA methylation patterns. *Genome Biol Evol* 5:1443–1456. <https://doi.org/10.1093/gbe/evt103>
- Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA (2010) GC₃ biology in corn, rice, sorghum and other grasses. *BMC Genomics* 11:308. <https://doi.org/10.1186/1471-2164-11-308>
- Tatarinova TV, Chekalin E, Nikolsky Y, et al (2016) Nucleotide diversity analysis highlights functionally important genomic regions. *Sci Rep* 6:35730. <https://doi.org/10.1038/srep35730>
- Triska M, Grocutt D, Southern J, et al (2013) cisExpress: motif detection in DNA sequences. *Bioinformatics* 29:2203–2205. <https://doi.org/10.1093/bioinformatics/btt366>
- Triska M, Solovyev V, Baranova A, et al (2017) Nucleotide patterns aiding in prediction of eukaryotic promoters. *PLoS One* 12:e0187243. <https://doi.org/10.1371/journal.pone.0187243>
- Troukhan M, Tatarinova T, Bouck J, et al (2009) Genome-wide discovery of cis-elements in promoter sequences using gene expression. *OMICS* 13:139–151. <https://doi.org/10.1089/omi.2008.0034>
- Wang P, Xia H, Zhang Y, et al (2015) Genome-wide high-resolution mapping of DNA methylation identifies epigenetic variation across embryo and endosperm in maize (*Zea mays*). *BMC Genomics* 16:21. <https://doi.org/10.1186/s12864-014-1204-7>
- Wen G (2017) A simple process of RNA-Sequence analyses by Hisat2, Htseq and DESeq2. In: *Proceedings of the 2017 International Conference on Biomedical Engineering and Bioinformatics - ICBEB 2017*. ACM Press, New York, New York, USA, pp 11–15
- West PT, Li Q, Ji L, et al (2014) Genomic distribution of H3K9me2 and DNA methylation in a maize genome. *PLoS One* 9:e105267. <https://doi.org/10.1371/journal.pone.0105267>
- Wingender E, Kel AE, Kel OV, et al (1997) TRANSFAC, TRRD and COMPEL: towards a federated database system on transcriptional regulation. *Nucleic Acids Res* 25:265–268. <https://doi.org/10.1093/nar/25.1.265>
- Yang H, Chang F, You C, et al (2015) Whole-genome DNA methylation patterns and complex associations with gene structure and expression during flower development in *Arabidopsis*. *Plant J* 81:268–281. <https://doi.org/10.1111/tbj.12726>
- Yan H, Bombarely A, Xu B, et al (2018) siRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass. *Biotechnol Biofuels* 11:208. <https://doi.org/10.1186/s13068-018-1202-0>
- Zemach A, Kim MY, Silva P, et al (2010) Local DNA hypomethylation activates genes in rice endosperm. *PNAS* 107:18729–18734. <https://doi.org/10.1073/pnas.1009695107>
- Zhang D, Wang Z, Wang N, et al (2014) Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. *PLoS One* 9:e96879. <https://doi.org/10.1371/journal.pone.0096879>
- Zhang Y-Y, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytol* 197:314–322. <https://doi.org/10.1111/nph.12010>

Figure Captions

Fig. 1 Structure of *EgDEF1* gene on oil palm genome scaffold (Accession No. NW_011551018.1). The *Karma* element and the associated transcript isoform *kDEF1* are displayed. Hypomethylation of the *Karma* element results in the expression of *kDEF1* in flowers of mantled clonal palms leading to the mantled fruit phenotype. *kDEF1* is not expressed in normal clonal palms.

Fig. 2 Circos plot of oil palm chromosomes displaying genes, repeats, mCG, mCHG, and mCHH patterns in the seedling. The window size is 1 Mb. Track order: (1) Mean gene density per window; (2) mean repeat density per window; mean methylation percentage per window for: (3) mCG; (4) mCHG and (5) mCHH.

Fig. 3 mCG, mCHG and mCHH profiles in various genomic regions of seedlings. The upstream region refers to 1,000 nt upstream of the start codon. Blue: mCG, orange: mCHG, green: mCHH.

Fig. 4 Profiles of average mCG, mCHG and mCHH across the gene, intragenic and intergenic TEs and their flanking 2 kb regions in seedlings. TSS: transcription start site, TTS: transcription termination site.

Fig. 5 GC₃ and DNA methylation gradients in the coding regions of genes in oil palm seedlings. **a** GC₃. **b** DNA methylation.

Fig. 6 Positional distribution and methylation patterns of transcription factor binding sites (TFBSs) in the oil palm genome. **a** promoter-specific TFBSs. **b** Non-positionally specific. **c** 5' UTR-specific TFBSs. **d** Box plot of mCG in promoter-specific TFBSs. **e** Box plot of mCG in 5' UTR-specific TFBSs. The red line shows the overall methylation of the region. Vertical lines show the position of the transcription start site.

Fig. 7 Influence of promoter-specific TFBS methylation on gene expression. The horizontal dotted line corresponds to the significance level cut-off. The window size is 8 nt.

Fig. 8 Pairwise comparison of mCG in seedling, normal, and mantled clones. The blue line represents lowess smoothed Z-score between seedling and normal, red line for the smoothed Z-score between seedling and mantled, the green line for the smoothed Z-score between normal and mantled. The window size is 10,000 nt. Red boxes with dashed lines indicate significant demethylation in regions of more than 10,000 nt in mantled clones ($|Z\text{-score}| > 3$, raw values are in the table ST5). Methylation differences between seedling, mantled and normal clonal palms can be witnessed in parts of chromosomes 1, 2, 3, and 5, which exhibited significant demethylation in mantled clones. Other chromosomes did not display significant Z-scores between seedling and mantled clones, or the significant demethylation occurred in a rather small region (10,000 nt) rather than the 1,000,000 nt region found in chromosome 1.

Fig. 9 Methylation of TE families in seedling (S), normal (N), and mantled (M) clones.

Fig. 10 mCHG patterns in the *EgDEF1* locus in mantled clones, normal clones, and seedling. Light-colored bars show statistically significant methylated cytosines (binom.test), dark-colored bars are for methylated/total reads ratio. Orange and green lines are Z-score values for seedling vs. mantled and seedling vs. normal data, respectively. Plots were smoothed out using ggplot2 built-in geom_smooth command under the loess method with a span of 0.2. The blue horizontal line specifies the Z-score threshold of 3 (refer secondary axis on the right). The dataset was filtered using a coverage threshold of 10. A significantly differentially methylated part of *Karma* transposon (from 3800 to 7000 nt) is shown as a red dotted box. The *Karma* sequence included in the *kDEF1* transcript is specified by gray vertical dashed lines within the boxed area. *EgDEF1* exons are shown at the bottom of the figure in gray boxes, while *Karma* is shown in orange.

Table Caption

Table 1. Genome-wide methylation in clones and seedlings (as calculated by Bismark (Krueger and Andrews 2011)). C: cytosine.