Biochimica et Biophysica Acta 1860 (2017) 149–156

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

Putting DNA methylation in context: from genomes to gene expression in plants☆

Chad E. Niederhuth, Robert J. Schmitz *

Department of Genetics, The University of Georgia, Athens, GA, 30602, USA

ARTICLE INFO

Article history:
Received 2 April 2016
Received in revised form 20 August 2016
Accepted 23 August 2016
Available online 30 August 2016

Keywords:
DNA methylation
Epigenetics
Epigenomics
Gene expression
Chromatin

ABSTRACT

Plant DNA methylation is its own language, interpreted by the cell to maintain silencing of transposons, facilitate chromatin structure, and to ensure proper expression of some genes. Just as in any language, context is important. Rather than being a simple "on-off switch", DNA methylation has a range of "meanings" dependent upon the underlying sequence and its location in the genome. Differences in the sequence context of individual sites are established, maintained, and interpreted by differing molecular pathways. Varying patterns of methylation within genes and surrounding sequences are associated with a continuous range of expression differences, from silencing to constitutive expression. These often-subtle differences have been pieced together from years of effort, but have taken off with the advent of methods for assessing methylation across entire genomes. Recognizing these patterns and identifying underlying causes is essential for understanding the function of DNA methylation and its systems-wide contribution to a range of processes in plant genomes. This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

1. Introduction

DNA is often called the "language of life", but it does not operate in isolation. Genetic information first passes through multiple layers of regulation before being expressed as phenotype. These layers of regulation range from protein-protein interactions and protein modifications, to modifications made to RNA and even DNA itself. Within all this is the epigenome; consisting of histones [1], histone modifications [2], non-coding RNAs, and the methylation of DNA [3]. These directly shape the structure of the genome, defining regions of euchromatin and heterochromatin, and helping to facilitate proper gene expression. DNA methylation, the addition of a methyl group to cytosine bases, is a key part of plant epigenomes. There is growing interest not only in the possible regulatory role of DNA methylation, but also its potential as an untapped source of diversity for plant improvement. Variants in DNA methylation, or "epialleles", can have phenotypic consequences and are in some cases "epigenetic", resulting in heritable phenotypic variation independent from any differences in DNA sequence [4]. The affects and associations of DNA methylation are context-dependent and recognizing this is critical to a proper understanding and utilization of DNA methylation.

Multiple techniques have been developed to assess DNA methylation [5]. However, the greatest advances has come with the ability to quantify methylation at the scale of whole genomes [6]. This first became possible with the advent of array-based methods [7]. Now with whole genome bisulfite sequencing (WGBS), the production of single-base resolution maps of DNA methylation across entire genomes is possible [8,9]. This is not without challenges in analysis and interpretation. For example, these techniques are typically applied to whole tissues consisting of mixed populations of plant cells. As a result, despite the fact that at any single base a cytosine is either methylated or not, these data represent an average of these mixed cell populations [10]. This requires careful consideration of how to determine methylation status and how to measure methylation levels. Other reviews have further discussion of the nuances of measuring and interpreting WGBS data [10].

Plants are a powerful system for studying DNA methylation and epigenetic phenomena. Although plant DNA methylation shares similarities and some of the same mechanisms as mammals [11,12], they have also evolved their own distinct pathways [12]. In plants, as in mammals, changes in DNA methylation are important during reproduction and in reproductive tissues, a topic that has been reviewed elsewhere [13]. However, whereas the epigenome of mammals is typically reset and re-established each generation [14]; in flowering plants there is not an analogous systematic genome-wide erasure of DNA methylation like there is in mammals. Instead, the parental state is typically reinforced and stably inherited [15]. As a consequence, the

☆ This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

* Corresponding author at: University of Georgia, 120 East Green Street, Athens, GA 30602.
E-mail address: schmitz@uga.edu (R.J. Schmitz).

http://dx.doi.org/10.1016/j.bbagrm.2016.08.009
1874-9399/

spontaneous formation of epialleles can be faithfully propagated between generations and there are multiple documented cases of epigenetic inheritance in plants [16–19]. This not only makes plants ideal for studying epigenetics, but also opens the possibility for utilizing or inducing epigenetic differences for purposes of plant breeding and improvement [20,21]. Much of what we now know about DNA methylation in plants comes from the study of one species, *Arabidopsis thaliana*, but this is rapidly expanding to other species [23–30]. This has opened up the field of comparative epigenomics, which applies the toolkit of comparative genomics to the study of the epigenome [22]. With it has come deeper insight into the evolution, mechanisms, and context-specific effects of DNA methylation. This review will focus on recent findings from epigenomics into the context-specific patterns of DNA methylation across the genome and within genes with discussion of sources of variation, their origins, potential functions, and potential applications.

2. Maintenance of DNA methylation is context dependent

At the most basic level, contextual differences exist between individual cytosines based on the sequence of neighboring bases. Plants methylate DNA in three different sequence contexts: dinucleotide CpG or CG (hereafter mCG) sites and trinucleotide CHG and CHH (H = A, T, or C) sites (hereafter mCHG and mCHH) [23,24]. This is in contrast to mammals where methylation is primarily found in the CpG context [25] and in specific cell types, the CpH context [26,27]. Additional base modifications to cytosines, such as 5-hydroxymethylation, also exist in mammals, but these have not been identified in plants despite extensive efforts [28]. This division of plant DNA methylation into various sequence contexts has a mechanistic basis, as differing enzymes and pathways are involved in the establishment and maintenance of DNA methylation for each context [12]. This makes the appropriate choice of methods and analysis critical for studying plant DNA methylation, as methods that cannot distinguish between sequence context obscure mechanistic and functional differences.

Perhaps the most familiar type of methylation is mCG, as it is found across eukaryotes [25,29]. Not only is mCG found across kingdoms, it is also the predominant type of methylation found in plants and animals [8,9,27]. This is due in part to the mechanisms that ensure its maintenance after DNA replication. CpG and CHG sites are symmetrical, as there is a mirrored cytosine on the opposing strand [23]. This is in contrast to CHH sites, which are asymmetrical [24]. The symmetrical nature of CpG sites is critical to the mechanism of its maintenance [30]. This is done by DNA (cytosine-5-)-METHYLTRANSFERASE 1 (DNMT1) in mammals [31] and its ortholog in plants METHYLTRANSFERASE 1 (MET1) [11,32,33]. In mammals, DNMT1 is associated with the replication machinery during DNA replication [34]. Due to the semiconservative replication of DNA, newly replicated DNA is hemi-methylated, consisting of one original methylated strand and one newly synthesized unmethylated strand. The mammalian Ubiquitin-like containing PHD and RING finger domains 1 (UHRF1) [30] and their plant orthologs, VARIANT IN METHYLATION 1 (VIM1-5) [35,36], recognize these hemi-methylated mCG sites and recruit DNMT1/MET1 to methylate the symmetrical site on the opposing strand. In this way, mCG is maintained across cell divisions. This pathway is essential to the continued maintenance of mCG, as *met1* mutants lose the majority of mCG in their genome [9,37]. Even after MET1 is restored by outcrossing the mutant to a wild-type plant and the mutant allele segregated away, CG methylation does not fully return [38,39].

CHG methylation, like mCG, is symmetrical [23], but it is maintained by a different pathway that involves CHROMOMETHYLASE 3 (CMT3) [40,41]. This enzyme is part of the plant-specific CMT family of methyltransferases, characterized by a conserved CHROMO domain [42,43]. The CHROMO domain, along with a BROMO ADJACENT HOMOLOGY (BAH) domain, facilitate the binding of the CMTs to di-methylated lysine 9 residues of histone 3 tails (H3K9me2) [44]. H3K9me2 is found primarily in regions of constitutive heterochromatin and CMT3 binds to H3K9me2, where it methylates CHG sites [41]. H3K9 methylation itself, is dependent upon mCHG [45]. The histone methyltransferases KRYPTONITE (KYP), Su(var)3-9 homologue 5 (SUVH5), and SUVH6 possess SRA domains that bind to methylated DNA, leading to the establishment of H3K9me2 [45–47]. This creates a self-reinforcing loop by which mCHG and H3K9me2 maintain their localization in the genome [45]. Indeed, loss of H3K9me2 in *kyp suvh5 suvh6* triple mutants leads to a loss of mCHG throughout those regions and mCHG-deficient mutants lead to reduced H3K9me2 in heterochromatin [48].

CHH methylation is asymmetrical and as such there is no mirrored cytosine on the opposite strand to serve as a guide for maintaining methylation after DNA replication [24]. Rather it is established *de novo* each round by one of two mechanisms. The first involves the activity of another member of the CMT family, CMT2 [49]. Like CMT3, it possesses a CHROMO and BAH domain and interacts with H3K9me2, leading to the methylation of CHH in H3K9me2 regions [44,49,50]. There is also some crossover between mCHH and mCHG. Although CMT2 preferentially methylates mCHH, it does to some extent methylate mCHG [50]. The bulk of CMT2 methylation is found within transposon bodies and within the pericentromere [49]. The ability of CMT2, CMT3, and MET1 to access and methylate heterochromatic regions is dependent in part on the chromatin remodeler DECREASED DNA METHYLATION 1 (DDM1) [49,51] and mutation of *DDM1* leads to a progressive loss of methylation in all three contexts over generations [52]. There also exists a third member of the family, CMT1, however, to date its function has not been identified [42]. Indeed, CMT1 is truncated in many *A. thaliana* accessions, suggesting that it may be non-functional [42].

Although all three contexts have their own distinct mechanisms, they are united by one common pathway, RNA-directed DNA Methylation (RdDM) [53]. In this pathway, 24-nucleotide (24-nt) short-interfering RNAs (siRNAs) [54] are dependent on the plant-specific POLYMERASE IV (POL IV) [55–58]. These siRNAs guide the *de novo* methyltransferase DOMAINS REARRANGED 2 (DRM2) to methylate target regions in all three contexts [59,60] (For an in-depth review see [12,53]). RdDM is in particular associated with high levels of mCHH and appears to target regions like the edges of transposons [49] and euchromatin-heterochromatin boundaries, like mCHH-islands in *Zea mays* [61,62].

There is also cross-talk between all these pathways. For example, in *met1* mutants, there is a redistribution of H3K9me2 and histone 3 lysine 27 tri-methylation (H3K27me3) to certain loci, leading to the accumulation of mCHG at new H3K9me2 sites [63]. Crosses between *met1* and wild-type plants can result in epigenetic "shock" in the F_1 progeny, inducing novel epialleles that do not exist in either parent [39]. The F_1 progeny of these crosses have widespread differences in small RNAs [39]. Many of the changes observed are in part due to demethylation of the seventh intron of *INCREASED IN BONSAI METHYLATION 1* (*IBM1*), a histone demethylase responsible for removing H3K9me2 from actively transcribed genes [39,64]. This causes a decrease in expression of *IBM1* and altered H3K9me2 distributions genome-wide, which subsequently alters targeting of CMT3 and CMT2 [39,64]. These results illustrate in part the complex interplay amongst the various DNA methylation pathways.

3. Genome-wide patterns of methylation in flowering plant genomes

Genomes are dynamic and vary in size [65,66], ploidy [67], gene content [68,69], repetitive elements [66] and more. DNA methylation is shaped by all these factors and vary considerably within species [70–72] and between species [29,73–76]. Within *A. thaliana*, methylation levels in all three contexts varied among accessions, differing as much as ~2x for mCG, ~6x for mCHG, and ~11x for mCHH [72]. Methylation levels differ by even greater degrees between plant species, with mCG varying ~3x, mCHG ~9x, and mCHH ~16x [74,75,77,78].

Underlying these genome-wide variations are dissimilarities in both the percentage of sites methylated and how highly those sites are methylated [75,76]. In Poaceae species (grasses), mCHH methylation is found primarily in small regions of highly methylated sites, whereas in some of the Fabaceae (legumes), like *Glycine max*, it is predominated by large regions of low methylation [75]. Genome-wide association analysis (GWAS) and quantitative trait loci (QTL) analysis have implied both cis and trans acting factors in explaining methylation differences [71,79] and many of these factors have now been identified.

The molecular pathways underlying DNA methylation are a major source of natural variation at all scales. This remains a source of variation within kingdoms as well. For instance, the roundworm *Caenorhabditis elegans* and the yeast *Saccharomyces cerevisiae* have no DNA methyltransferases in their genome and also no cytosine DNA methylation [80,81]. Although methylation has been found in all plant genomes to date, certain pathways have been found to predominate and others to be defective. Several examples come from the Brassicaceae. In *Brassica rapa* a mutation in NRPD1, the largest subunit of Pol IV, leads to reduced production of Pol IV-dependent siRNAs [82]. Across the Brassicaceae methylomes sequenced to date, per-site levels of mCHG are typically lower than in other species, with the lowest being in the species *Eutrema salsugineum* and *Conringia planisiliqua* [75,83]. Analysis of these species genomes shows that they have both independently lost CMT3, while phylogenetics studies show that CMT3 in other Brassicaceae are under relaxed selection, accumulating non-synonymous mutations at a higher rate [83,84]. GWAS within *A. thaliana* revealed natural variation in methylation levels of transposons was due to a premature stop codon in CMT2 in many wild accessions [72]. Similarly, in the Poaceae, CMT2 is absent in *Z. mays* [49], where the majority of mCHH is due to the activity of RdDM [85]. So while plants have multiple unique methylation enzymes and pathways, these continue to evolve in a lineage specific manner, leading to increasing epigenomic diversity.

There is a strong tie between DNA methylation and the silencing of repetitive DNA, which has a strong impact on the epigenome [86–88]. Methylation levels can vary within repeats and transposons themselves depending on the particular transposon family [89] and potentially also its age [90]. Although polyploidy can and does contribute to genome size variation, it is the expansion and reduction of repetitive elements that are the primary drivers [65,66]. As the number of repeats and transposons in a genome increase, so does the intergenic spaces and overall genome size. It is expected that overall methylation levels will vary both with genome size and repeat content. Indeed, among the Brassicaceae, much of the variation of DNA methylation observed between species is driven by differences in repetitive elements, this is especially true in conserved syntenic regions [74]. When viewed across the angiosperms, total mC (methylation in all contexts) is correlated to increasing genome size [91]. Broken down methylation contexts, both mCG and mCHG are correlated to increases in genome size and the number of repeats in the genome, surprisingly though mCHH does not [75]. This could be driven in part by differences in mCHH in the Poaceae, which have low mCHH levels, yet typically large genomes.

4. Functionality of nonCG methylated genes and hypothetical origins

Within plant species there are genes that are consistently methylated in all cytosine sequences (Fig. 1A) [71]. In *A. thaliana* many of these genes are targets of RdDM and these show tissue-specific expression, being reactivated in pollen [71]. These genes share common functions based on gene ontology (GO); being enriched for hydrolases, cell wall modification, and translation machinery. These processes are essential for pollen tube tip growth, which suggests a possible developmental regulation of these RdDM-targeted loci [71]. This reactivation of silenced genes coincides with the activation of many transposons from the vegetative nucleus [71]. In *A. thaliana*, *DDM1* is not expressed in the pollen vegetative nucleus and this is associated with the expression of transposons as well as the production of epigenetically activated

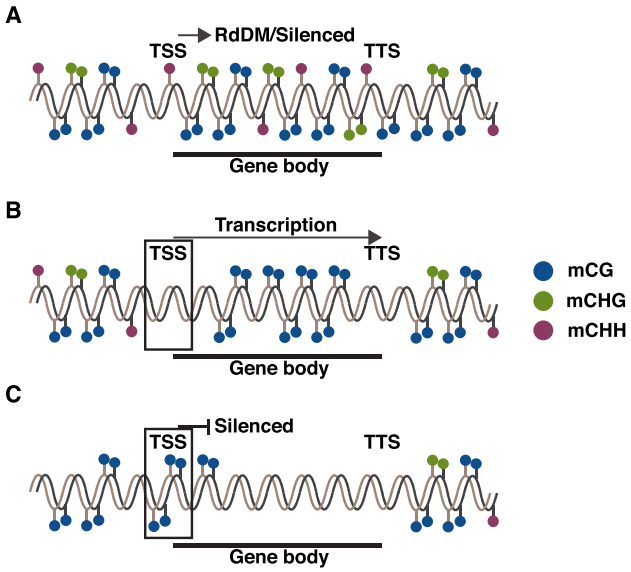

Fig. 1. A) Genes with nonCG methylation (mCHG and mCHH) within gene bodies are typically targets of RNA-directed DNA methylation (RdDM) and silenced. **B)** Gene-body methylated genes are expressed and methylated in the CG context only. They are characterized by a depletion of mCG around the transcriptional start site (TSS), increasing mCG through the gene-body, and a depletion of mCG over the transcriptional termination site (TTS). **C)** Methylation of CG over the TSS is associated with silencing of expression.

small RNAs (easiRNAs) that are capable of moving to the germline nuclei [15]. It is thought that these easiRNAs are one mechanism by which epigenome states are reinforced each generation in flowering plants [15]. Some of these silenced genes are also enriched for defense responses [79]. Methylation and gene expression changes do occur in newly dividing cells upon bacterial infection [92,93]. It is possible that these methylated genes are in fact pseudogenes and accumulate mutations due to the mutagenic effects of DNA methylation [94]. However, many of these silenced genes show no evidence of accumulating loss-of-function mutations in *A. thaliana* [71], which further supports a functional role for many of these genes.

Although within species comparisons imply a functionality of nonCG methylation of genes in development or as part of environmental responses, nonCG methylated states are not conserved in orthologous genes between species [74,75]. This suggests that it largely arises *de novo* within interbreeding populations, diverging with increasing genetic distance. Indeed, clustering of *A. thaliana* accessions based on differentially methylated regions largely reconstitutes the genetic and geographic relationship of these lines [71]. Many possible hypotheses could explain this. For example, nonCG methylation might target lineage-specific or orphan genes as in the case of the *Qua-Quine Starch* (*QQS*) gene in *A. thaliana* [95]. However, this is insufficient to explain the number of nonCG methylated genes in some species, which can constitute up to a third of all genes [75].

In highly conserved regions between *A. thaliana*, *A. lyrata*, and *C. rubella*, transposon gains and losses alone explain much of methylation differences between species [74]. In many species there is an enrichment of transposons upstream and downstream of nonCG methylated genes [75]. At a genome-wide level, correlations are also observed for methylation levels within the coding sequence (CDS) and genome size for nonCG methylation [75,76]. Spreading of methylation from transposons into gene bodies is known to result in silencing [96], with specific examples being the case of *CmWIP1* in *C. melo* [97]. In *Z. mays*, spreading of heterochromatic states from transposons into neighboring regions and genes is limited to specific transposon families [89]. Presence-absence variation of transposons resulted in gene-expression differences between lines [89]. Between species there is a correlation of increasing CDS methylation levels

with increasing repeat content in genes and for many species, an enrichment of transposons upstream and downstream of nonCG genes [75]. GWAS in *A. thaliana* indicates that ~35% of differentially methylated regions are associated with local genetic variants and this is likely an underestimate [71] as QTL analysis of differentially methylated regions in *G. max* indicates a much higher proportion segregates with the parental genotype [79]. Two examples from *A. thaliana* demonstrate how copy-number variation and gene structure can lead to methylation and silencing. There are three *PHOSPHORIBOSYLANTHRANILATE ISOMERASE* (*PAI1-3*) genes in most *A. thaliana* accessions and in some accessions, these genes are highly methylated and silenced [98,99]. It was discovered that in lines where *PAI1-3* are silenced, there is a fourth copy, *PAI4*, that is tandemly duplicated on the opposite strand of *PAI1* resulting in a *PAI1-PAI4* inverted repeat, which is sufficient to induce silencing of all *PAI* genes in the genome [98–100]. In another example, a truncated and rearranged copy of FOLT2 in certain A. thaliana accessions leads to silencing and methylation of *FOLT1* [101]. Methylation and silencing of *FOLT1* is stable and inherited even after the inducing *FOLT2* copy is segregated away [101].

Despite the lack of conservation of nonCG methylation between species, there is a degree of similarity at a functional level for nonCG genes when gene ontology is taken into account, as the same GO terms are enriched even across distantly related species [75]. These include processes such as proteolysis, cell death, and defense responses and could reflect the strict regulatory control that such processes are under. Many species also show enrichment for processes like photosynthesis, electron transport chain, and primary metabolism [75]. At first glance, this is unexpected, as these are essential functions and not expected to be silenced. Closer investigation of these loci shows that typically these nuclear genes are orthologous to mitochondrial and chloroplast genes and are likely insertions from organellar genomes into the nuclear genome [75]. Indeed, the movement of DNA from plant organelles into the nucleus is a constant and ongoing process and bear the marks of accumulated mutations due to methylation [102–105]. No single hypothesis for how genes become the targets of nonCG methylation and silencing exists. Rather a combination of multiple mechanisms is most likely involved in the origins of these silenced loci.

5. Gene-body methylation in flowering plants

DNA methylation is typically thought of as a repressive chromatin modification, however, this is not always the case. In both plants and animals, methylation of gene bodies in only the mCG context is associated with higher gene expression and more specifically, constitutively expressed genes (Fig. 1B) [29,37,106–108]. These gene-body methylated (gbM) genes characterized by depletion of mCG in and around the transcriptional start site (TSS) and transcriptional termination sites (TTS) [107]. The pattern of methylation appears to be a critical aspect, as methylation across the TSS by mCG alone is sufficient for transcriptional silencing (Fig. 1C) [75].

The widespread nature of gbM across species [29] and its unexpected association with gene expression begs the question as to the functional role of gbM and the origin of gbM genes [109]. The methylation status of gbM genes is often conserved across orthologs between species, as in the case of *O. sativa* and *B. distachyon* [110]. These results having been generalized more recently across a greater range of species [74–76]. Although conservation is observed, it breaks down over evolutionary distance. Within a species, there is natural variation in which genes are gbM, as demonstrated in *A. thaliana* [72,111]. Both gains and losses of gbM are observed between closely related species like *A. lyrata* and *A. thaliana* or *O. sativa* and *B. distachyon* [75,110]. In one accession of *M. guttatus*, ~60.7% of genes are gbM [75], whereas at the other extreme there is a depletion of gbM genes in many Brassicaceae [75,83]. Outside of the angiosperms, gbM is known to be absent in more basal plant species such as *Marchantia polymorpha* [76] and *Selaginella moellendorffii* [29]. In other non-angiosperms, such as ferns, both mCG and mCHG are found within gene-bodies and do not show typical patterns of mCG (a depletion at the TSS and TTS) [76,84]. These results suggest that within plants, gbM as defined in this review article, is specific to angiosperms [84]. Within angiosperms, gbM has been completely lost at least twice independently in two Brassicaceae: *E. salsugineum* and *C. planisiliqua* [75,83]. Comparison of gene expression of gbM loci from *A. thaliana* and orthologs in *E. salsugineum* indicates no overt effects on gene expression even though one group has gbM and the other group does not [83]. Many hypotheses have been put forward as to the possible functions of gbM. These include the exclusion of the histone variant H2A.Z, alternative splicing, the suppression of antisense transcription, and as a dampener on gene expression [107,112–114]. Analysis of the transcriptome, histone variants, and histone modifications in *E. salsugineum* failed to support any of these hypotheses [83]. Collectively, these observations indicate that gbM is dispensable over evolutionary time.

A second line of evidence that gbM is dispensable comes from studies of epigenetic recombinant inbred lines (epiRILs) [38,83]. The epiRILs were generated by crossing *A. thaliana met1* mutants, which lose nearly all mCG in the genome, to genetically identical wild-type plants [38]. In the F_2 generation, plants homozygous for wild-type *MET1* were retained and selfed for eight generations. This created a population of genetically identical plants, which varied in their methylation status, being a mosaic of methylated regions from the wild-type parent and unmethylated regions from the *met1* mutant [38]. This epiRIL population enabled the effects of gbM and its loss to be tested in genes that previously possessed it [83]. As in the case of *E. salsugineum*, there was no evidence that loss of gbM resulted in changes in gene expression, alternative splicing, or in the distribution of H2A.Z [83]. Results from gbM loss in *E. salsugineum* and *C. planisiliqua*, as well as the epiRILs, indicate no functional consequences of gbM, at least for current hypotheses, however it is possible there are yet undiscovered functions of gbM.

6. Origins of gene-body methylation in flowering plants

Given the evidence showing gbM is dispensable, how are additional associations explained? Beyond gene expression, many other structural and evolutionary features of gbM genes have been identified. Structurally, gbM genes in plants are typically longer and have more exons when compared to non-gbM genes [83,110,115]. They also have been reported to be more slowly evolving, having a lower ratio of nonsynonymous to synonymous mutations (*dN*/*dS*) [83,110,115]. The slower rate of evolution, however, seems to break down with increasing evolutionary distance [75]. These may be accidental associations, however, united by a common link of transcription. For example, in plants, longer genes are typically more highly expressed [116] and expression levels are known to contribute to evolutionary rates [117,118]. Associations between gbM and these features could arise as a result that these genes are more actively transcribed.

The independent loss of gbM and *cmt3* in both *E. salsugineum* and *C. planisiliqua* supports the hypothesis that they are connected (Fig. 2). One possible explanation suggest the involvement of IBM1, a histone demethylase that removes H3K9me2 from actively transcribed genes [119]. Mutations in *ibm1* result in H3K9me2 accumulation specifically in gbM loci [119]. This leads not only to increased mCHG through the recruitment of CMT3 within these gene bodies, but also to increased mCG and mCHH through yet unknown means (Fig. 2) [83]. That IBM1 is necessary for the active removal of H3K9me2 in gbM genes, suggests that H3K9me2 is found at these loci at low levels in wild-type plants. Once mCG is established in the gene-body, as long as it does not impede normal gene function, it would be faithfully maintained by MET1 (Fig. 2) [83]. In the absence of mechanisms to re-establish new mCG, gbM would decay slowly due to the fidelity of maintenance dependent on MET1 and with the epimutation rate [17,83,120]. This would explain why loss of gbM has not been observed in

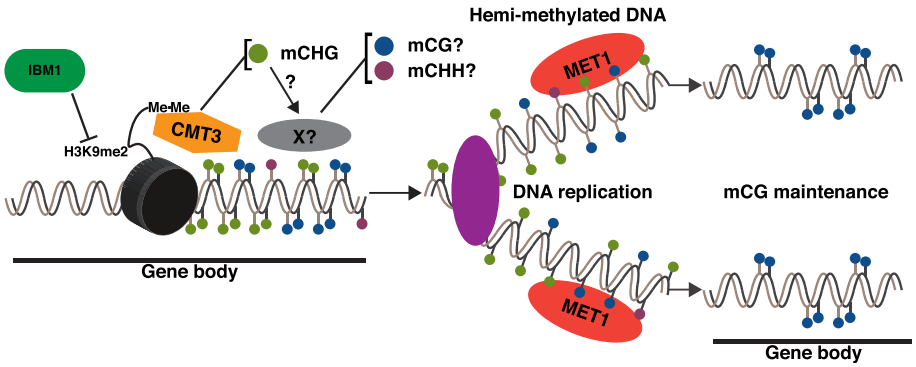

Fig. 2. Proposed model for the establishment of gene-body methylation. IBM1 actively removes H3K9me2 from transcribed genes, indicating that H3K9me2 at some level is incorporated in these genes. In *ibm1* mutants there is an accumulation of CG, CHG, and CHH methylation in transcribed genes. CMT3 is associated with H3K9me2 and methylates CHG sites. Unknown factors (X?) mediate methylation of CG and CHH sites. After DNA replication, MET1 maintains CG methylation in transcribed genes, while CHG and CHH methylation is lost.

A. thaliana cmt3 mutants and poses a challenge for proving such a hypothesis as it may require many generations to observe. The example of gbM in angiosperms and its association with CMT3 demonstrates the power of comparative epigenomics for hypothesis generation and for hypothesis testing.

7. Methylation of regulatory regions and gene expression

Methylation of regulatory elements of genes also affects transcription factor (TF) binding and gene expression. In *A. thaliana*, the genome-wide cis-element binding (cistrome) has been assessed for 529 different TFs using DNA affinity purification sequencing (DAP-seq) [121]. Modified versions of this technique were also used to assess the impact of DNA methylation on TF occupancy. The majority (72%) of TFs were inhibited by methylation, whereas 24% showed no inhibition. Interestingly, 4.3% of TFs preferentially bound methylated DNA, further illustrating the context-specific effects of DNA methylation. A fascinating example of this involves the regulation of a DNA glycosylase, REPRESSOR OF SILENCING 1 (ROS1), which leads to demethylation of DNA [122,123]. In this example, methylation of a neighboring TE in the promoter region of *ROS1* promotes gene expression, whereas loss of methylation suppresses it. It is proposed that this serves as a "rheostat", fine tuning the expression of *ROS1* to variations in the methylation level of the genome [122,123].

Transcription factor occupancy or gene expression itself may drive changes in DNA methylation [124]. One of the best examples to date in plants examined the relationship between DNA methylation and gene expression of *Oryza sativa* and *A. thaliana* under phosphate starvation [125]. By sampling from multiple time points, Secco and colleagues (2015) found that gene expression changes preceded changes in methylation in upstream or downstream regions in *O. sativa*. Most of the methylation differences observed in *O. sativa* were associated with TEs, in particular MINITURE INVERTED REPEATS (MITEs). Upon resupply of phosphate, gene expression returned to levels found in plants that were not subject to phosphate-starvation. Methylation profiles lagged behind gene-expression changes, being more similar to phosphate-starved samples even after 31 days of recovery [125]. Meanwhile, few methylation differences where associated with phosphate-starvation in *A. thaliana* [125]. These results strongly suggest that in certain cases, it is changes in gene expression that drive methylation differences in regulatory regions, rather than the other way around.

In *Z. mays* the presence of high levels of mCHH upstream and downstream of genes has been observed and this is associated with genes that are more highly expressed [61]. These regions have been termed "CHH islands" and they were originally found to be enriched among MITE transposons [61] and other terminal inverted repeat (TIR) sequences [62]. Methylated CHH (mCHH) islands are primarily the result of RdDM pathways, as they are lost in the *mediator of paramutation 1* (*mop1*) and *mop3* mutants, known components of RdDM in *Z. mays* [61,62]. Like so many aspects of DNA methylation, their association with gene expression begs the question of which came first. In *mop1* and *mop3* mutants, loss of mCHH islands has no affect on neighboring gene expression and presence-absence variation between accessions was found indicating that mCHH islands do not drive gene expression [62]. Furthermore, although mCHH islands could be identified across species [75], many showed no association with differences in gene expression. Rather, the primary factor in determining their abundance appears to be the presence of repetitive elements upstream or downstream of genes, as there is a correlation between the two [61,62,75]. Indeed, differences in the distribution of repeats near genes explains the depletion and enrichment of mCHH islands between species [75]. The presence of higher mCG and mCHG levels upstream or downstream of mCHH islands in the direction away from the gene and changes in histone modifications indicates that mCHH islands may mark a boundary between euchromatin and localized heterochromatin in plant genomes [62]. The examples of *ROS1* and mCHH islands, shows that there is much we still do not know about DNA methylation and its impact on the regulatory regions of genes.

8. Epialleles as a source of phenotypic variation in nature and in agriculture

As plants do not reset their epigenome each generation in a manner similar to mammals, changes to methylated states that occur in the germline can be passed on to offspring, resulting in epigenetic variation. One classic example is the *peloric* mutant in *Linaria vulgaris*. In wild-type plants, flowers have dorsoventral asymmetry, however, in the *peloric* mutant, the flowers display a radially symmetrical floral phenotype [126]. These mutants have a storied history, having been described in 1749 by Carl Linnaeus. 250 year later, the basis of this mutation was identified as the hypermethylation and silencing of *Lcyc*, the *L. vulgaris* ortholog of the *cycloidea* gene in *Antirrhinum majus* [126]. Many other examples have been identified since: the *Colorless non-ripening* (*Cnr*) locus in *S. lycopersicum* (tomato) [127,128], *CmWIP1* in *C. melo* (melon) [98], and the *PAI* [99,100,129], *FOLT1*/*FOLT2* [101], and *QQS* genes in *A. thaliana* [95,130]. Epialleles like *Cnr* and *CmWIP1* directly impact traits of agricultural production and importance. The *Cnr* epialle in tomato leads to silencing of a *SQUAMOSA* promoter binding protein–like essential to fruit ripening [129]. This results in an inhibition of ripening, producing instead a colorless fruit with a loss of cell adhesion and unfavorable characteristics [127,128]. The epiallele at *CmWIP1*, as previously mentioned, results in spreading of DNA methylation from a neighboring transposon [98]. *CmWIP1* is involved in flower sex determination in melons and in the silenced epiallele, produces female flowers. With

microarrays and whole-genome bisulfite sequencing, many more such epialleles have and are being identified [16,17,79,131].

The ability of epialleles to impact gene expression and subsequently agriculturally important phenotypes has sparked interest in their use for agriculture [21]. An excellent example is the *MANTLED* locus of *Elaeis guineensis* (African oil palm) [132]. Micropropagation of *E. guineensis* leads to a high frequency of plants producing sterile floral organs, greatly reducing yield. In an epigenome-wide association study, this was mapped to loss of methylation at a transposable element called *KARMA* [132]. *KARMA*, is located in the intron of the gene *DEFICIENS* and loss of methylation leads to mis-splicing of *DEFICIENS* and improper flower formation [132]. Having identified the underlying epiallele, it will be possible to test micropropagated plants for this epiallele prior to planting. Epialleles could also potentially be used in breeding programs, either as a source of novel phenotypic variation [38,133–135] or by using epialleles as molecular markers for creating genetic maps and mapping quantitative traits [134]. Indeed, modern breeding methods like genomic selection are already being adapted for use of epigenetic markers [136].

9. Conclusion

As part of the epigenome, DNA methylation is an additional layer through which the language of the genome is interpreted. However, context is important and meanings can change with subtle variations in patterns of the epigenome. DNA methylation is not a simple on-off switch. In plants, DNA can be methylated at three sequence contexts: CG, CHG, and CHH. These are established and maintained by distinct molecular mechanisms, and in the cases of mCHG and mCHH, processes that are unique to plants. The effect of methylation on gene expression is highly dependent upon the type of methylation as well as the pattern of that methylation within or outside of the gene. Decades of work have gone into understanding the nature of DNA methylation, which is now being accelerated by whole genome approaches and a flood of new data. While much progress has been made there remains much that is unknown, in particular the effects of methylation outside of gene bodies upon gene expression and how targeting of certain genes by DNA methylation originates. The full extent of epigenetic variation is also not understood, despite its potential to impact phenotypic variation and traits of agricultural importance. Wider studies that take from multiple populations, species, and mutants will be needed for a deeper understanding of DNA methylation and its integration into how it affects gene expression.

Conflict of interest

The authors report no conflict of interest.

Acknowledgements

This work was supported by the National Science Foundation (NSF) (MCB – 1339194, IOS - 1546867) and the National Institutes of Health (R00GM100000) to R.J.S. C.E.N was supported by a NSF postdoctoral fellowship (IOS – 1402183).

References

[1] P.B. Talbert, S. Henikoff, Histone variants–ancient wrap artists of the epigenome, Nat. Rev. Mol. Cell Biol. 11 (4) (2010) 264–275.
[2] K. Zhang, et al., Distinctive core histone post-translational modification patterns in Arabidopsis thaliana, PLoS One 2 (11) (2007) e1210.
[3] E.J. Finnegan, et al., DNA Methylation in Plants, Annu. Rev. Plant Physiol. Plant Mol. Biol. 49 (1998) 223–247.
[4] E.J. Richards, Inherited epigenetic variation–revisiting soft inheritance, Nat. Rev. Genet. 7 (5) (2006) 395–401.
[5] L. Shen, R.A. Waterland, Methods of DNA methylation analysis, Curr. Opin. Clin. Nutr. Metab. Care 10 (5) (2007) 576–581.
[6] R. Lister, J.R. Ecker, Finding the fifth base: genome-wide sequencing of cytosine methylation, Genome Res. 19 (6) (2009) 959–966.
[7] A. Schumacher, et al., Microarray-based DNA methylation profiling: technology and applications, Nucleic Acids Res. 34 (2) (2006) 528–542.
[8] S.J. Cokus, et al., Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning, Nature 452 (7184) (2008) 215–219.
[9] R. Lister, et al., Highly integrated single-base resolution maps of the epigenome in Arabidopsis, Cell 133 (3) (2008) 523–536.
[10] M.D. Schultz, R.J. Schmitz, J.R. Ecker, 'Leveling' the playing field for analyses of single-base resolution DNA methylomes, Trends Genet. 28 (12) (2012) 583–585.
[11] E.J. Finnegan, E.S. Dennis, Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana, Nucleic Acids Res. 21 (10) (1993) 2383–2388.
[12] J.A. Law, S.E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, Nat. Rev. Genet. 11 (3) (2010) 204–220.
[13] T. Kawashima, F. Berger, Epigenetic reprogramming in plant sexual reproduction, Nat. Rev. Genet. 15 (9) (2014) 613–624.
[14] S. Seisenberger, et al., Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 368 (1609) (2013) 20110330.
[15] R.K. Slotkin, et al., Epigenetic reprogramming and small RNA silencing of transposable elements in pollen, Cell 136 (3) (2009) 461–472.
[16] C. Becker, et al., Spontaneous epigenetic variation in the Arabidopsis thaliana methylome, Nature 480 (7376) (2011) 245–249.
[17] R.J. Schmitz, et al., Transgenerational epigenetic instability is a source of novel methylation variants, Science 334 (6054) (2011) 369–373.
[18] E. Heard, R.A. Martienssen, Transgenerational epigenetic inheritance: myths and mechanisms, Cell 157 (1) (2014) 95–109.
[19] J. Hagmann, et al., Century-scale methylome stability in a recently diverged Arabidopsis thaliana lineage, PLoS Genet. 11 (1) (2015) e1004920.
[20] E. Jaligot, A. Rival, Applying Epigenetics in Plant Breeding: Balancing Genome Stability and Phenotypic Plasticity, 2015 159–192.
[21] L. Ji, D.A. Neumann, R.J. Schmitz, Crop Epigenomics: Identifying, Unlocking, and Harnessing Cryptic Variation in Crop Genomes, Mol. Plant 8 (6) (2015) 860–870.
[22] X. Zhong, Comparative epigenomics: a powerful tool to understand the evolution of DNA methylation, New Phytol. 22 (1) (2015) 76–80.
[23] Y. Gruenbaum, Sequence specificity of methylation in higher plant DNA, Nature 292 (5826) (1981) 860–862.
[24] P. Meyer, I. Niedenhof, M. ten Lohuis, Evidence for cytosine methylation of non-symmetrical sequences in transgenic Petunia hybrida, EMBO J. 13 (9) (1994) 2084–2088.
[25] A.P. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (2002) 6–21.
[26] B.H. Ramsahoye, et al., Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, Proc. Natl. Acad. Sci. 97 (10) (1994) 5237–5242.
[27] R. Lister, et al., Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (7271) (2009) 315–322.
[28] R.M. Erdmann, et al., 5-hydroxymethylcytosine is not present in appreciable quantities in Arabidopsis DNA, G3 (Bethesda) **5** (1) (2014) 1–8.
[29] A. Zemach, et al., Genome-wide evolutionary analysis of eukaryotic DNA methylation, Science 328 (5980) (2010) 916–919.
[30] M. Bostick, et al., UHRF1 plays a role in maintaining DNA methylation in mammalian cells, Science 317 (5845) (2007) 1760–1764.
[31] T. Bestor, et al., Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases, J. Mol. Biol. 203 (4) (1988) 971–983.
[32] E.J. Finnegan, W.J. Peacock, E.S. Dennis, Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development, Proc. Natl. Acad. Sci. U. S. A. 93 (16) (1996) 8449–8454.
[33] M.J. Ronemus, et al., Demethylation-induced developmental pleiotropy in Arabidopsis, Science 273 (5275) (1996) 654–657.
[34] P.M. Vertino, et al., DNMT1 is a component of a multiprotein DNA replication complex, Cell Cycle 1 (6) (2002) 416–423.
[35] H.R. Woo, et al., VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization, Genes Dev. 21 (3) (2007) 267–277.
[36] H.R. Woo, T.A. Dittmer, E.J. Richards, Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis, PLoS Genet. 4 (8) (2008) e1000156.
[37] X. Zhang, et al., Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis, Cell 126 (6) (2006) 1189–1201.
[38] J. Reinders, et al., Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes, Genes Dev. 23 (8) (2009) 939–950.
[39] M. Rigal, et al., Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in Arabidopsis thaliana F1 epihybrids, Proc. Natl. Acad. Sci. U. S. A. 113 (14) (2016) E2083–E2092.
[40] A.M. Lindroth, et al., Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation, Science 292 (5524) (2001) 2077–2080.
[41] J.P. Jackson, et al., Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase, Nature 416 (6880) (2002) 556–560.
[42] S. Henikoff, L. Comai, A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis, Genetics 149 (1) (1998) 307–318.
[43] A. Pavlopoulou, S. Kossida, Plant cytosine-5 DNA methyltransferases: structure, function, and molecular evolution, Genomics 90 (4) (2007) 530–541.

[44] J. Du, et al., Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants, Cell 151 (1) (2012) 167–180.
[45] L.M. Johnson, X.F. Cao, S.E. Jacobsen, Interplay between two epigenetic marks: DNA methylation and histone H3 lysine 9 methylation, Curr. Biol. 12 (16) (2002) 1360–1367.
[46] M.L. Ebbs, L. Bartee, J. Bender, H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases, Mol. Cell. Biol. 25 (23) (2005) 10507–10515.
[47] M.L. Ebbs, J. Bender, Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase, Plant Cell 18 (5) (2006) 1166–1176.
[48] H. Stroud, et al., Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome, Cell 152 (1-2) (2013) 352–364.
[49] A. Zemach, et al., The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin, Cell 153 (1) (2013) 193–205.
[50] H. Stroud, et al., Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis, Nat. Struct. Mol. Biol. 21 (1) (2014) 64–72.
[51] A. Vongs, et al., Arabidopsis thaliana DNA methylation mutants, Science 260 (5116) (1993) 1926–1928.
[52] T. Kakutani, J.A. Jeddeloh, E.J. Richards, Characterization of an Arabidopsis thaliana DNA hypomethylation mutant, Nucleic Acids Res. 23 (1) (1995) 130–137.
[53] M.A. Matzke, R.A. Mosher, RNA-directed DNA methylation: an epigenetic pathway of increasing complexity, Nat. Rev. Genet. 15 (6) (2014) 394–408.
[54] S.W. Chan, et al., RNA silencing genes control de novo DNA methylation, Science 303 (5662) (2004) 1336.
[55] A.J. Herr, et al., RNA polymerase IV directs silencing of endogenous DNA, Science 308 (5718) (2005) 118–120.
[56] Y. Onodera, et al., Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation, Cell 120 (5) (2005) 613–622.
[57] J. Luo, B.D. Hall, A multistep process gave rise to RNA polymerase IV of land plants, J. Mol. Evol. 64 (1) (2007) 101–112.
[58] T.S. Ream, et al., Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II, Mol. Cell 33 (2) (2009) 192–203.
[59] X. Cao, S.E. Jacobsen, Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes, Proc. Natl. Acad. Sci. U. S. A. 99 (Suppl. 4) (2002) 16491–16498.
[60] X. Cao, S.E. Jacobsen, Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing, Curr. Biol. 12 (13) (2002) 1138–1144.
[61] J.I. Gent, et al., CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize, Genome Res. 23 (4) (2013) 628–637.
[62] Q. Li, et al., RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome, Proc. Natl. Acad. Sci. U. S. A. 112 (47) (2015) 1428–1433.
[63] A. Deleris, et al., Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana, PLoS Genet. 8 (11) (2012) e1003062.
[64] M. Rigal, et al., DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns, EMBO J. 31 (13) (2012) 2981–2993.
[65] J.L. Bennetzen, J. Ma, K.M. Devos, Mechanisms of recent genome size variation in flowering plants, Ann. Bot. 95 (1) (2005) 127–132.
[66] J.L. Bennetzen, H. Wang, The contributions of transposable elements to the structure, function, and evolution of plant genomes, Annu. Rev. Plant Biol. 65 (2014) 505–530.
[67] P.S. Soltis, et al., Polyploidy and genome evolution in plants, Curr. Opin. Genet. Dev. 35 (2015) 119–125.
[68] J. Sebat, et al., Large-scale copy number polymorphism in the human genome, Science 305 (5683) (2004) 525–528.
[69] R.A. Swanson-Wagner, et al., Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor, Genome Res. 20 (12) (2010) 1689–1699.
[70] S.R. Eichten, et al., Epigenetic and genetic influences on DNA methylation variation in maize populations, Plant Cell 25 (8) (2013) 2783–2797.
[71] R.J. Schmitz, et al., Patterns of population epigenomic diversity, Nature 495 (7440) (2013) 193–198.
[72] M.J. Dubin, et al., DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation, Elife 4 (2015) e05255.
[73] S. Feng, et al., Conservation and divergence of methylation patterning in plants and animals, Proc. Natl. Acad. Sci. U. S. A. 107 (19) (2010) 8689–8694.
[74] D.K. Seymour, et al., Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization, PLoS Genet. 10 (11) (2014) e1004785.
[75] C.E. Niederhuth, et al., Widespread natural variation of DNA methylation within angiosperms, bioRxiv (2016), http://dx.doi.org/10.1101/045880.
[76] S. Takuno, J.-H. Ran, B.S. Gaut, Evolutionary patterns of genic DNA methylation vary across land plants, Nat. Plants 2 (2) (2016) 15222.
[77] J.M. Colicchio, et al., DNA methylation and gene expression in Mimulus guttatus, BMC Genomics 16 (2015) 507.
[78] H. Wang, et al., CG gene body DNA methylation changes and evolution of duplicated genes in cassava, Proc. Natl. Acad. Sci. U. S. A. 112 (44) (2015) 13729–13734.
[79] R.J. Schmitz, et al., Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population, Genome Res. 23 (10) (2013) 1663–1674.
[80] J.H. Proffitt, et al., 5-Methylcytosine Is Not Detectable in Saccharomyces cerevisiae DNA, Mol. Cell. Biol. 4 (5) (1984) 985–988.
[81] S. Yi, Birds do it, bees do it, worms and ciliates do it too: DNA methylation from unexpected corners of the tree of life, Genome Biol. 13 (10) (2012) 174.
[82] Y. Huang, T. Kendall, R.A. Mosher, Pol IV-Dependent siRNA Production is Reduced in Brassica rapa, Biology (Basel) 2 (4) (2013) 1210–1223.
[83] A.J. Bewick, et al., On the origin and evolutionary consequences of gene body DNA methylation, Proc. Natl. Acad. Sci. U. S. A. 113 (32) (2016) 9111–9116.
[84] A.J. Bewick, et al., The evolution of chromomethyltransferases and gene body DNA methylation in plants, bioRxiv (2016), http://dx.doi.org/10.1101/054924.
[85] Q. Li, et al., Genetic perturbation of the maize methylome, Plant Cell 26 (12) (2014) 4602–4616.
[86] V.L. Chandler, V. Walbot, DNA modification of a maize transposable element correlates with loss of activity, Proc. Natl. Acad. Sci. U. S. A. 83 (6) (1986) 1767–1771.
[87] P.S. Chomet, S. Wessler, S.L. Dellaporta, Inactivation of the Maize Transposable Element Activator (Ac) Is Associated with Its DNA Modification, EMBO J. 6 (2) (1987) 295–302.
[88] P.D. Rabinowicz, et al., Differential methylation of genes and repeats in land plants, Genome Res. 15 (10) (2005) 1431–1440.
[89] S.R. Eichten, et al., Spreading of heterochromatin is limited to specific families of maize retrotransposons, PLoS Genet. 8 (12) (2012) e1003127.
[90] F. Maumus, H. Quesneville, Ancestral repeats have shaped epigenome and genome composition for millions of years in Arabidopsis thaliana, Nat. Commun. 5 (2014) 4104.
[91] C. Alonso, et al., Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms, Front. Genet. 6 (2015) 4.
[92] R.H. Dowen, et al., Widespread dynamic DNA methylation in response to biotic stress, Proc. Natl. Acad. Sci. U. S. A. 109 (32) (2012) E2183–E2191.
[93] A. Yu, et al., Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense, Proc. Natl. Acad. Sci. U. S. A. 110 (6) (2013) 2389–2394.
[94] J.C. Shen, W.M. Rideout 3rd, P.A. Jones, The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA, Nucleic Acids Res. 22 (6) (1994) 972–976.
[95] A.B. Silveira, et al., Extensive natural epigenetic variation at a de novo originated gene, PLoS Genet. 9 (4) (2013) e1003437.
[96] C. Weil, R. Martienssen, Epigenetic interactions between transposons and genes: lessons from plants, Curr. Opin. Genet. Dev. 18 (2) (2008) 188–192.
[97] A. Martin, et al., A transposon-induced epigenetic change leads to sex determination in melon, Nature 461 (7267) (2009) 1135–1138.
[98] J. Bender, G.R. Fink, Epigenetic Control of an Endogenous Gene Family Is Revealed by a Novel Blue Fluorescent Mutant of Arabidopsis, Cell 83 (5) (1995) 725–734.
[99] S. Melquist, B. Luff, J. Bender, Arabidopsis PAI gene arrangements, cytosine methylation and expression, Genetics 153 (1) (1999) 401–413.
[100] B. Luff, L. Pawlowski, J. Bender, An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis, Mol. Cell 3 (4) (1999) 505–511.
[101] S. Durand, et al., Rapid establishment of genetic incompatibility through natural epigenetic variation, Curr. Biol. 22 (4) (2012) 326–331.
[102] S. Stegemann, et al., High-frequency gene transfer from the chloroplast genome to the nucleus, Proc. Natl. Acad. Sci. U. S. A. 100 (15) (2003) 8828–8833.
[103] C.Y. Huang, et al., Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes, Plant Physiol. 138 (3) (2005) 1723–1733.
[104] T. Kleine, U.G. Maier, D. Leister, DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis, Annu. Rev. Plant Biol. 60 (2009) 115–138.
[105] L.M. Roark, et al., Recent and frequent insertions of chloroplast DNA into maize nuclear chromosomes, Cytogenet. Genome Res. 129 (1-3) (2010) 17–23.
[106] R.K. Tran, et al., DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes, Curr. Biol. 15 (2) (2005) 154–159.
[107] D. Zilberman, et al., Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription, Nat. Genet. 39 (1) (2007) 61–69.
[108] T. Baubec, et al., Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation, Nature 520 (7546) (2015) 243–247.
[109] F.K. Teixeira, V. Colot, Gene body DNA methylation in plants: a means to an end or an end to a means? EMBO J. 28 (8) (2009) 997–998.
[110] S. Takuno, B.S. Gaut, Gene body methylation is conserved between plant orthologs and is of evolutionary consequence, Proc. Natl. Acad. Sci. U. S. A. 110 (5) (2013) 1797–1802.
[111] T. Kawakatsu, et al., Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions, Cell 166 (2) (2016) 492–505.
[112] D. Zilberman, et al., Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks, Nature 456 (7218) (2008) 125–129.
[113] D. Coleman-Derr, D. Zilberman, Deposition of histone variant H2A.Z within gene bodies regulates responsive genes, PLoS Genet. 8 (10) (2012) e1002988.
[114] M. Regulski, et al., The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA, Genome Res. 23 (10) (2013) 1651–1662.
[115] S. Takuno, B.S. Gaut, Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly, Mol. Biol. Evol. 29 (1) (2012) 219–227.
[116] X.Y. Ren, et al., In plants, highly expressed genes are the least compact, Trends Genet. 22 (10) (2006) 528–532.
[117] C. Pal, B. Papp, L.D. Hurst, Highly expressed genes in yeast evolve slowly, Genetics 158 (2) (2001) 927–931.
[118] L. Yang, B.S. Gaut, Factors that contribute to variation in evolutionary rate among Arabidopsis genes, Mol. Biol. Evol. 28 (8) (2011) 2359–2369.

[119] A. Miura, et al., An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites, EMBO J. 28 (8) (2009) 1078–1086.
[120] A. van der Graaf, et al., Rate, spectrum, and evolutionary dynamics of spontaneous epimutations, Proc. Natl. Acad. Sci. U. S. A. 112 (21) (2015) 6676–6681.
[121] R.C. O'Malley, et al., Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape, Cell 165 (5) (2016) 1280–1292.
[122] M. Lei, et al., Regulatory link between DNA methylation and active demethylation in Arabidopsis, Proc. Natl. Acad. Sci. U. S. A. 112 (11) (2015) 3553–3557.
[123] B.P. Williams, et al., Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat, PLoS Genet. 11 (3) (2015) e1005142.
[124] A. Feldmann, et al., Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions, PLoS Genet. 9 (12) (2013) e1003994.
[125] D. Secco, et al., Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements, Elife 4 (2015).
[126] P. Cubas, C. Vincent, E. Coen, An epigenetic mutation responsible for natural variation in floral symmetry, Nature 401 (6749) (1999) 157–161.
[127] A.J. Thompson, et al., Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant, Plant Physiol. 120 (2) (1999) 383–390.
[128] K. Manning, et al., A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening, Nat. Genet. 38 (8) (2006) 948–952.
[129] J. Bender, G.R. Fink, Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis, Cell 83 (5) (1995) 725–734.
[130] L. Li, et al., Identification of the novel protein QQS as a component of the starch metabolic network in Arabidopsis leaves, Plant J. 58 (3) (2009) 485–498.
[131] S.R. Eichten, et al., Heritable Epigenetic Variation among Maize Inbreds, PLoS Genet. 7 (11) (2011) e1002372.
[132] M. Ong-Abdullah, et al., Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm, Nature 525 (2015) 533–537.
[133] F. Johannes, et al., Assessing the impact of transgenerational epigenetic variation on complex traits, PLoS Genet. 5 (6) (2009) e1000530.
[134] S. Cortijo, et al., Mapping the epigenetic basis of complex traits, Science 343 (6175) (2014) 1145–1148.
[135] M. Dapp, et al., Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids, Nat. Plants 1 (7) (2015) 15092.
[136] Y. Hu, et al., Prediction of Plant Height in Arabidopsis thaliana Using DNA Methylation Data, Genetics 201 (2) (2015) 779–793.