Epigenetic modifications in plants: an evolutionary perspective
Suhua Feng¹ and Steven E Jacobsen¹,²

Plant genomes are modified by an array of epigenetic marks that help regulate plant growth and reproduction. Although plants share many epigenetic features with animals and fungi, some epigenetic marks are unique to plants. In different organisms, the same epigenetic mark can play different roles and/or similar functions can be carried out by different epigenetic marks. Furthermore, while the enzymatic systems responsible for generating or eliminating epigenetic marks are often conserved, there are also cases where they are quite divergent between plants and other organisms. DNA methylation and methylation of histone tails on the lysine 4, 9, and 27 positions are among the best characterized epigenetic marks in both plants and animals. Recent studies have greatly enhanced our knowledge about the pattern of these marks in various genomes and provided insights into how they are established and maintained and how they function. This review focuses on the conservation and divergence of the pathways that mediate these four types of epigenetic marks.

Introduction
Plants show a high degree of developmental plasticity, partly owing to their sessile way of life and the need to cope with a frequently changing environment. Recent studies show that epigenetic pathways (e.g. DNA methylation, histone variants and modifications, positioning of nucleosomes, and small RNA) are important components of plant growth and reproduction regulation [1⁰]. Multiple aspects of plant development, including flowering time, gametogenesis, stress response, light signaling, and morphological change are modulated directly or indirectly by epigenetic marks. Plants have acquired complex systems to regulate the epigenetic marks on their genomic DNA, some of which are conserved from other organisms such as animals and insects, and some of which are specific for plants.

In the genomes of higher plants, 5-methyl cytosine methylation is widely found, bearing the important function of defense against activation and movement of transposable elements and expression regulation of certain developmental genes. DNA methylation is conserved in many other eukaryotic organisms, albeit with clear divergence in the methylation enzyme systems and functions [2⁰,3⁰,4⁰]. Other common epigenetic marks consist of modifications on histone tails. Unlike DNA methylation that invariably takes place at the carbon-5 position of cytosine residues, different histones (H2A, H2B, H3, and H4) can be covalently modified at different positions (mostly lysine and arginine residues) by different chemical marks (methylation, acetylation, ubiquitination, phosphorylation, biotinylation, and ADP-ribosylation) [5,6]. Different histone marks have different functions, and even the same histone mark can have different functions in different organisms. Three types of histone methylation in plants, histone H3K4 mono/di/tri-methylation (H3K4me1, H3K4me2, and H3K4me3), histone H3K27 tri-methylation (H3K27me3) and histone H3K9 di-methylation (H3K9me2), are the most well-studied representatives to consider the evolution of plant epigenetic modification systems.

H3K4 mono/di/tri-methylation
Lysines are capable of accepting three methyl groups, meaning that a given lysine residue on histone can be mono-methylated, di-methylated, or tri-methylated. In the case of H3K4 in Arabidopsis, these three methylated forms are all detected by mass spectrometry analysis [6]. Genome-wide profiling of Arabidopsis H3K4me1, H3K4me2, and H3K4me3 by ChIP-chip demonstrated that they exist exclusively in genes and promoters (~2/3 of all genes) and are essentially absent from euchromatic regions where transposons and repetitive DNA reside (Table 1 and Figure 1), consistent with the notion that H3K4 methylation marks the active chromatin [7⁰]. Within genes, the distributions of the three H3K4 marks are different—H3K4me1 is enriched in the body of the genes with depletion on both ends of the genes, while H3K4me2 and H3K4me3 are enriched in the promoter and 5’-end of genes with H3K4me3 being further upstream of H3K4me2 (Figure 1). Furthermore, only H3K4me3 is associated with active transcription while H3K4me1 and H3K4me2 are not well correlated with transcription [7⁰].

Similar distribution patterns of the three types of H3K4 methylation have been reported in other organisms,
including rice, yeast, and human [8–12]. This suggests that the mechanism for H3K4 methylation is highly conserved. Yeast (Saccharomyces cerevisiae) has a single H3K4 methyltransferase, termed SET1, which has a highly conserved SET (Su(var)3-9, Enhancer-of-zeste, and Trithorax) domain (C24 150 amino acids). SET1 forms a complex called Complex Proteins Associated with Set 1 (COMPASS), which can mediate mono-methylation, di-methylation, and tri-methylation of H3K4 [13]. In Drosophila, H3K4 methylation is mediated by homologs of yeast SET1, the Trithorax group (TrxG) proteins, which were first identified genetically as the counteractors of the Polycomb group (PcG) proteins in controlling the expression of Homeotic (HOX) genes (see below) [14]. Another Drosophila H3K4 methyltransferase that also contains a SET domain is Absent, Small, or Homeotic Disc 1 (Ash1), which was suggested to be the main enzyme responsible for di-methylating H3K4 [15]. Mammals methylate H3K4 by COMPASS-like complexes that contain various TRX-family proteins, as well as Ash1 [16,17] (Table 1). SET-domain H3K4 methyltransferases are also found in Arabidopsis, in the form of five Arabidopsis TRX proteins, ARABIDOPSIS HOMOLOG OF TRITHORAX 1 (ATX1) to ATX5; as well as seven ARABIDOPSIS TRITHORAX-RELATED (ATXR) proteins, ATXR1 to ATXR7; and seven Ash1 homologs, ASH1 HOMOLOG 1 (ASHH1) to ASHH4, and ASH1-RELATED 1 (ASHR1) to ASHR3 [18] (Table 1). Among them, ATX1 and ATX2 are the best studied. These two proteins appear to play quite divergent roles despite their highly similar protein sequence (C24 65% identical). For instance, they regulate the transcription of two largely non-overlapping sets of genes [19,20]. It has been shown at some loci that H3K4me3 is mediated by ATX1 whereas H3K4me2 is mediated by ATX2 [20,21]. Moreover, ASHH2, also known as EARLY FLOWERING IN SHORT DAYS (EFS) and SET DOMAIN GROUP 8 (SDG8), is a dual function histone methyltransferase for both H3K4 and H3K36 [22]. Most recently, ATXR3, also known as SDG2, has been demonstrated to be the major H3K4 tri-methyltransferase in Arabidopsis [23,24].

Table 1

<table>
<thead>
<tr>
<th>Epigenetic marks and corresponding players in Arabidopsis and human</th>
<th>Mark</th>
<th>Modifying system</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histone methylations</td>
<td>H3K4me1/me2/me3</td>
<td>Arabidopsis: ATX proteins, ATXR3, and ASHH2</td>
<td>Genes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human: TrxG proteins, Ash1</td>
<td>Genes</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Arabidopsis: PcG proteins (no PRC1 components though)</td>
<td>Human:</td>
<td>Genes</td>
</tr>
<tr>
<td>H3K9me2</td>
<td>Arabidopsis: KYP, SuvH5/6</td>
<td></td>
<td>Heterochromatin</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Human: Suvh39h proteins</td>
<td></td>
<td>Heterochromatin</td>
</tr>
<tr>
<td>DNA methylations</td>
<td>CG</td>
<td>Arabidopsis: MET1</td>
<td>Genes, TEs, and repeats</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human: Dnmt1</td>
<td>Everywhere except CpG islands</td>
</tr>
<tr>
<td>CHG</td>
<td>Arabidopsis: CMT3</td>
<td>TEs and repeats</td>
<td></td>
</tr>
<tr>
<td>CHH</td>
<td>Arabidopsis: DRM2</td>
<td>TEs and repeats</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1

Schematic representation of the distribution of selected epigenetic marks in the Arabidopsis genome. The genome of Arabidopsis can be divided into two portions, pericentromeric heterochromatin and the euchromatic chromosome arms. Pericentromeric heterochromatin containing abundant transposons and silenced genes is characterized by large regions of high levels of H3K9me2 and DNA methylation in all three sequence contexts. Transposons found in the euchromatin also contain H3K9me2 and three types of DNA methylation, but are present as small patches of heterochromatin limited to the length of the transposon. Some genes in euchromatin, which do not have DNA methylation, are repressed by H3K27me3. Expressed genes often have methylated H3K4, with tri-methylation and di-methylation in the promoter and 5′-end and mono-methylation in the transcribed region. Genes with modest levels of transcription tend to have gene body CG methylation. Histone methylations are illustrated on the top and DNA methylations on the bottom. m: methylated.
Overall, it is clear that H3K4 methylation systems are evolutionarily ancient and that plants and other eukaryotes probably share a common ancestral mechanism. However, it is also clear that certain TRX-related proteins have acquired other functions in plants. For example, ATXR5 and ATXR6 are two methyltransferases for H3K27 mono-methylation (a repressive mark for silencing transposons) and their SET domains are quite diverged from the ones present in SET1 homologs. Interestingly, this methylation also appears to regulate DNA replication in heterochromatin [25*].

H3K27 tri-methylation

Another abundant histone modification in Arabidopsis is H3K27 tri-methylation [5,6]. H3K27me3 has been extensively studied in Arabidopsis, as well as in many other organisms, as a major repressive mark for gene expression. Several well-known Arabidopsis developmental genes, including flower timing gene FLOWERING LOCUS C (FLC), floral organ patterning gene AGAMOUS (AG), homeobox gene SHOOT MERISTEMLESS (STM), and two imprinted genes MEDEA (MEA) and PHERES1 (PHE1), are epigenetically silenced by H3K27me3 [26,27]. Recently, many more H3K27me3 target genes (~4400) have been revealed by whole-genome ChIP-chip analysis in Arabidopsis [28**] (Table 1). These genes are enriched for transcription factors, supporting an important role of H3K27me3 in plant development. The expression levels of the H3K27me3 modified genes are very low and often exhibit a high degree of tissue specificity (with the majority of them only expressed in one or a few tissues), which suggests repression of these genes by H3K27me3 is alleviated only in the place where their expression is needed. Interestingly, H3K27me3-modified regions in Arabidopsis are generally limited to the length of a single gene and two or more adjacent genes controlled by the same patch of H3K27me3 is very rarely seen [28**] (Figure 1). This is in contrast to the long-range spreading and very large patches of H3K27me3 that is common in Drosophila and mammals [29–31].

This difference might be attributed to the divergence of H3K27 tri-methylation systems in plants versus other organisms. All organisms that have H3K27me3 contain Polycomb group proteins (PeG). Like TrxG proteins, PeG proteins were also first identified through genetic analysis in Drosophila owing to their effect (repressive, as opposed to the activation role of TrxG) on HOX genes [32]. Several protein complexes are formed by PeG, namely Polycomb Repressive Complex 1 (PRC1), PRC2, and PhoRC [33]. PRC2 contains a key subunit called Enhancer of zeste, E(z), which is a SET domain histone methyltransferase specific for H3K27 tri-methylation. Arabidopsis has three E(z) homologs, CURLY LEAF (CLF), MEA, and SWINGER (SWN), as well as homologs for each of the other subunits of PRC2 [26,27] (Table 1). It therefore seems likely that PRC2 was present in the last common ancestor of plants and animals. However, the Arabidopsis genome does not seem to encode components for PRC1 or PhoRC, suggesting that these two complexes are either lost in the plant lineage or have evolved independently in animals. Considering the implicated role of PRC1 in recognizing and assisting in the spread of H3K27me3 [34], the absence of PRC1 in plants might explain why the average length of H3Kk27me3-modified regions in Arabidopsis is a few kilobases as opposed to hundreds of kilobases seen in Drosophila and mammals [28**]. Nonetheless, plants must have systems that recognize the H3K27me3 mark in order for it affects gene expression. Several PRC1-like activities have been reported in Arabidopsis. A plant chromodomain protein, LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), has been found to bind H3K27me3 in vitro (through its chromodomain) and colocalize with H3K27me3 genomewide in planta [35*,36*], which is analogous to the function of the chromodomain protein Polycomb (Pc) in the PRC1 complex. Other proteins that have been reported to substitute PRC1 functions in Arabidopsis include VERNALIZATION 1 (VRN1) [37], EMBRYONIC FLOWER 1 (EMF1) [38], and AtRING1a and AtRING1b [39,40]. This functional diversification may lead to different outcomes for PeG-regulated genes in plants, and this area deserves further study.

H3K9 di-methylation

In eukaryotes, heterochromatin is distinguished from euchromatin in that it is densely compacted, transcriptionally inactive, and contains methylated DNA, histones with repressive marks, and deacetylated histones. In animals and fungi, the formation of heterochromatin is partly dependent on the methylation of H3K9 and the interaction between methylated H3K9 and Heterochromatin Protein 1 (HP1) that contains a chromodomain for binding methylated histones [41–44]. In general, H3K9 tri-methylation is a mark of heterochromatin (e.g. in Neurospora crassa and mammals); however, H3K9me3 of Arabidopsis is typically localized in euchromatin [36**]. Instead, the pre-dominant mark for heterochromatin in Arabidopsis is H3K9 di-methylation [45] (Table 1 and Figure 1). Recent genomic profiling studies using ChIP-chip in Arabidopsis show that H3K9me2 is highly enriched in pericentromeric heterochromatin as large and uninterrupted blocks and also exists in euchromatic repeats and transposons as small patches that cover the respective repeat/transposon unit, consistent with the notion that it labels the silenced chromatin [46**] (Figure 1). H3K9me2-modified regions are tightly correlated with regions in the Arabidopsis genome that contain CHG (where H is A, C, or T) methylation (a type of DNA methylation almost exclusively found in higher plants; see below) [46**].

The enzymes for methylating H3K9, the Su(var)3-9 family proteins, were the first histone lysine methyltrans-
ferases reported [47]. The Su(var)3-9 locus, again, was initially identified through Drosophila genetics when searching for suppressors of position-effect variegation (PEV) [48]. The PEV suppressor screen also turned up Su(var)2-5 that encodes HP1. This was the initial indication that H3K9 methylation and HP1 probably functioned in the same pathway, which was later confirmed in many organisms [14]. Su(var)3-9 homologs have been shown in many organisms to play roles in heterochromatin formation and gene silencing, and a few examples include cryptic loci regulator 4 (Clr4) in yeast (S. pombe) and Suv39h proteins in mammals [42,49] (Table 1). Arabidopsis also has multiple Su(var)3-9 homologs called the SUVH proteins, and among them, KRYPTONITE (KYP, also known as SUVH4) is a mono-methyltransferase and di-methyltransferase for H3K9 that is required for the presence of H3K9me2 in heterochromatin [50,51]. Two other SUVH proteins, SUVH5 and SUVH6, also methylate H3K9 [50,52,53] (Table 1). Although plants are similar to animals and fungi in that they utilize Su(var)3-9 proteins for H3K9 methylation, plants have a very unique mechanism for the maintenance and function of H3K9 methylation. As mentioned above, the HP1 homolog in Arabidopsis, LHP1, recognizes and co-localizes with H3K27 tri-methylation but not H3K9 methylation [35**]. Arabidopsis H3K9me2 is instead bound by a different chromodomain-containing protein called CHROMOMETHYLASE 3 (CMT3), which is a maintenance DNA methyltransferase for CHG sites. Consistently, loss of KYP leads to reduction in both H3K9me2 and CHG methylation levels, suggesting H3K9me2 controls CHG methylation [51,54]. Moreover, the SRA (SET and RING-Associated) domain of KYP has been shown to bind DNA with methylated CHG sites, suggesting that DNA methylation recruits histone methyltransferase [55]. These findings support a self-reinforcing feedback model between KYP and CMT3 that efficiently maintains H3K9 methylation and CHG methylation in heterochromatic regions. A similar mechanism has been discovered in Neurospora, where H3K9me3 directs DNA methylation. However, it differs from the system in plants in two key aspects. First it requires HP1 to act as an adapter between the histone H3K9 methyltransferase Defective in Methylation 5 (DIM-5) and the DNA methyltransferase DIM-2. Second, histone methylation is strictly upstream of DNA methylation and there is therefore no feedback loop as for Arabidopsis [56].

DNA methylation

Cytosine methylation is a common modification found in genomes of plants, animals, and fungi. For instance, model organisms used for biological studies such as Arabidopsis, Neurospora, human, mouse, rice, and zebrafish contain abundant amount of methylated cytosines. However, DNA methylation has been curiously lost in some other well-studied model organisms including Caenorhabditis elegans, Drosophila, baker’s yeast, and fission yeast [57**]. Cytosines are methylated in a variety of DNA sequences contexts, but mechanistically can be classified broadly into three contexts, CG, CHG (H = A, T, C) and CHH [2**]. Because methylated cytosines behave the same way as unmethylated cytosines during standard DNA sequencing reactions, genome sequencing projects do not provide DNA methylation information. This can be overcome by sodium bisulfite treatment that converts unmethylated cytosines to uracils but does not alter methylated cytosines [58**,59**]. One complication in assaying DNA methylation is that it is highly variable even within the same cell type, which means that a particular cytosine position can show a different methylation status from one cell to another. Therefore, multiple (usually >10) sequenced clones covering the same cytosine are needed to obtain an overall picture of the methylation status for a given cytosine position (on either the Watson or Crick strand). With the recent advancement in high-throughput sequencing, high coverage methylation maps of eukaryotic genomes have started to emerge [58**,59**].

Arabidopsis has perhaps the most extensively characterized methylome of any organism. Owing to its small genome size and important role as a model system, Arabidopsis has become the first organism where a whole-genome tiling array analysis of DNA methylation and a whole-genome single-nucleotide resolution DNA methylation map were published [58**,59**,60]. Two general patterns of DNA methylation are evident in the Arabidopsis genome. The first is high levels of methylation in all three cytosine contexts (CG, CHG, and CHH) on transposable elements (TEs) and other repetitive DNA, which are mostly found in pericentromeric heterochromatic regions but also exist in small patches between genes in the euchromatic arms. The second is methylation in the transcribed region or body of genes (excluded from both ends and assuming a bell-like shape with a slight bias toward the 3'-half). This gene body methylation is found in ~1/3 of all protein-coding genes and takes place exclusively in the CG context [58**,59**,60–63] (Table 1 and Figure 1). Functionally, these two types of methylation play two very different roles. Methylation on TEs and repeats represses the transcription of these DNAAs as a genome defense mechanism against selfish DNA. On the contrary, gene body methylation somewhat positively correlates with gene transcription levels, with the highest methylation level observed in genes with moderately high transcription [4**,60,62]. Genes with tri-methylated H3K27 generally do not have DNA methylation, indicating the anti-correlation of these two epigenetic marks [28**] (Figure 1).

Recent whole-genome methylation analysis of a variety of eukaryotic organisms allows an examination of these two general patterns of methylation from an evolutionary perspective [3**,4**,57**]. Preferential methylation of
TEs and repeats has long been considered to be ancient, and perhaps the primary reason for the existence of DNA methylation [64]. However, TE methylation was not found in a number of invertebrate animals, including insects such as honeybee and silk moth, sea anemone, and sea squirt; yet, these organisms show a clear preference for methylation within gene bodies [3*]. Most plants and fungi, on the contrary, clearly preferentially methylate their TEs and repeats. Gene body methylation is found in both animals and plants, but not in fungi, suggesting that this may be an ancient methylation pattern that was subsequently lost in fungi. Vertebrate animals show both gene body methylation and transposon methylation. However, vertebrate genomes are so highly CG methylated (~85%) that it is somewhat difficult to assess whether TEs and repeats are preferentially methylated over the rest of the genome [3*,4*]. These considerations suggest that gene body methylation is likely to be at least as ancient as the TE and repeat methylation, both of which would be predicted to be present in the last common ancestor of animals, fungi, and plants.

Some algal species show unique patterns of methylation. Chlorella methylation patterns basically mimic those of vertebrates both at genome-wide levels and within the body of genes. Volvox on the contrary has very low levels of methylation overall, but like higher plants shows methylation of both genes and repeats [4*]. Chlamydomonas is unusual because it displays preferential methylation of genes in all three sequence contexts, instead of just in a CG context as found in most other organism; and it displays transposon methylation in a CG only context, instead of in all sequence contexts [3*].

The function of gene body methylation is unclear since its loss in methylation mutants has only subtle effects on overall levels of gene expression [60,62]. However, the recent finding that methylation is much more prevalent on exons than on introns suggests that methylation may contribute to exon definition or regulate alternative splicing [3*,65,66*]. Although it is an attractive hypothesis that methylation may regulate splicing, experimental support for this idea is generally lacking, and understanding the function of gene body methylation is an important future endeavor.

The eukaryotic cytosine methyltransferase enzymes that methylate DNA are homologous to bacterial restriction modification methyltransferases, revealing their very ancient origin [67]. The activity of DNA methyltransferases can be broadly classified into that which establishes methylation on previously unmethylated DNA (de novo methylation) and that which maintains preexisting methylation (maintenance methylation). De novo methylation in mammal and plants are mostly carried out through the DNA (cytosine-5)-methyltransferase 3 (Dnmt3) class of enzymes, called DOMAIN REARRANGED METHYLTRANSFERASE 2 (DRM2) in plants (Table 1) [2*]. However, the mechanism by which these enzymes are targeted is very different. Dnmt3 class enzymes in mammals are targeted in large part by binding to histone H3 tails that are unmethylated at lysine 4, which could help explain why mammalian genomes are heavily methylated except at the CpG-rich promoters of genes that are high in H3K4 methylation [1*,68]. In addition, DRM2 de novo methylation activity is targeted to DNA by small interfering RNAs (siRNAs) in a very complex pathway termed RNA-directed DNA methylation (RdDM) [2*]. Although an RdDM-like pathway does not exist in mammals, a close one is the PIWI-associated RNA (piRNA) pathway that guides Dnmt3 activity in mouse germ cells [69].

As proposed by Arthur Riggs more than 35 years ago [70], maintenance methylation relies at least partly on the symmetry of the CG and CHG sites. For the maintenance of CG methylation, the mechanism appears to be highly conserved, at least in plants and vertebrates. The GC maintenance DNA methyltransferase is high conserved and found in all plants animals and fungi. For instance Dnmt1 of mammals and METHYLTRANSFERASE 1 (MET1) (Table 1) from Arabidopsis appear to be orthologous and are similar in function. Mutations in both lead to dramatic losses of CG DNA methylation (in Arabidopsis this loss is complete) [58*,59*,67]. In addition, both function with a conserved cofactor called Ubiquitin-like Containing PHD and RING Finger Domains 1 (UHRF1) in mammals and VARIATION IN METHYLATION (VIM) in Arabidopsis [2*]. This cofactor contains an SRA domain that binds methylated DNA, and the SRA of UHRF1 has been demonstrated to recognize hemimethylated DNA, the physiological substrate for Dnmt1 that is produced at DNA replication foci [55,71–73].

CHG is also a symmetrical site, but the mechanism by which this methylation is maintained differs in plants and other organisms. As discussed above, plants maintain high levels of CHG methylation through a self-reinforcing feed-forward loop between the CMT3 DNA methyltransferase and the KYP H3K9 methyltransferase. Mammals have low amounts of CHG methylation, except in ES cells where it is clearly detectable along with CHH methylation. However, this CHG methylation is ‘asymmetrical’, meaning that CHG sites are usually only methylated only on one strand, and moreover, CHG and CHH methylation are of roughly similar levels, suggesting that they may be maintained by the same mechanism, probably through Dnmt3 activity [74*] (Table 1). CMT family methyltransferases are found only in plants and algae, where abundant CHG methylation is observed [3*,4*]. Fungal DIM-2 has similar functions to CMT3 in that it is also guided by histone H3K9 methylation (see above); however, DIM-2 does not appear to be specific for
CHG but rather methylates cytosines in all sequence contexts without any preference [56]. Both CMT and DIM-2 homologs are related to Dnmt1 but form a distinct group by themselves [57**].

CHH methylation is also maintained, but owing to its asymmetric nature it has been long thought that this type of methylation is probably persistently targeted by de novo DNA methylation systems. Consistent with this idea, CHH methylation in mammals is probably dependent on Dnmt3 class enzymes, and CHH methylation in Arabidopsis is dependent on DRM2 and RdDM [2**,74**].

Finally, although Dnmt1 enzymes are clearly the key maintenance methyltransferases for CG sites as discussed, some organisms such as algae and silk moth only have Dnmt1 but not Dnmt3 [3**,57**,67]. Interestingly, CHG and CHH methylation are readily detectable in the green algae Chlamydomonas [3**]. Together, these findings suggest that Dnmt1 might have assumed a de novo methylation function and/or adopted activity toward non-CG sites in some organisms.

**Conclusions and perspectives**

The conservation and divergence of multiple epigenetic modification pathways in plants and other eukaryotic organisms have started to be revealed by genetic and genomic studies of a variety of organisms (Table 1). A general theme that emerges is that the epigenetic marks and the mechanisms that establish these marks are frequently ancient and conserved, but the precise details of how these marks function within genomes is often divergent. This functional divergence is probably owing to the evolutionary forces that have adapted these epigenetic mechanisms to the needs of the specific organism. Another important factor not discussed in this review is the targeted erasure of epigenetic modifications, exemplified by histone and DNA demethylases. These activities, which show their own conservation and divergence of mechanism [1**,75**], act in opposition to the establishment and maintenance mechanisms to shape dynamic epigenomic landscapes.

**Acknowledgements**

We thank Xiaoyu Zhang for critical reading and commenting on the manuscript. Research in the Jacobsen laboratory is supported by the National Institutes of Health (GM60398) and the National Science Foundation Genome Research Program (#0701745). Suhua Feng is a Special Fellow of the Leukemia & Lymphoma Society. Steve Jacobsen is an investigator of the Howard Hughes Medical Institute.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


This article reviews recent findings on how the epigenomes of plants and animals are reprogrammed during reproductive growth and embryonic development. In particular, decreases of DNA methylation in nurse cells of gametophytes and seeds (in plants) and genome-wide waves of DNA demethylation in primordial germ cells and zygotes (in animals) are discussed.


This article reviews the mechanisms of DNA methylation in plants and animals. It covers the current knowledge on many aspects of DNA methylation, such as DNA methyltransferases, de novo methylation, maintenance methylation, methylation in reproductive cells, and DNA demethylation.


See annotation to [4**].


This study, along with [3**], used genomic bisulfite sequencing approaches to survey DNA methylation in multiple eukaryotic organisms and revealed the conservation and divergence of DNA methylation patterns in genes and in transposons.


A whole-genome ChiP-chip analysis of Arabidopsis H3K4 methylation marks. It showed that H3K4 is methylated exclusively in genes with different distribution patterns for mono-methylation, di-methylation, and tri-methylation.


20. Saleh A, Alvarez-Venegas R, Yilmaz M, Le Q, Hou G, Saddler M, Al-Abdallat A, Xia Y, Lu G, Ladunga I et al.: The highly similar Arabidopsis histone methyltransferase activities (ATX1 for tri-methylation of H3K4 and ATX2 for di-methylation). This paper analyzed trithorax proteins ATX1 and ATX2 and showed that these proteins carry different histone methyltransferase activities (ATX1 for tri-methylation of H3K4 and ATX2 for di-methylation). Their target genes are also largely non-overlapping.


46. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE:
   • Genome-wide association of histone H3 lysine nine
     methylation with CHG DNA methylation in Arabidopsis
A whole-genome ChiP-chip analysis of Arabidopsis H3K9 di-methylation.
H3K9me2 was found to be highly correlated with CHG methylation in
Arabidopsis. The major target of H3K9me2 is heterochromatin and it also
exists on transposons in the euchromatic arms.

47. Rea S, Eisenhaber F, O’Carroll D, Strahl BD, Sun ZW, Schmid M,
   Opravil S, Mechtler K, Ponting CP, Allis CD et al.: Regulation of
   chromatin structure by site-specific histone H3

   for position-effect variegation in Drosophila melanogaster.

49. Peters AH, O’Carroll D, Scherthan H, Mechtler K, Sauer S,
   Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A
   et al.: Loss of the suc39h histone methylestersases impairs
   mammalian heterochromatin and genome stability. Cell 2001,
   107:323–337.

50. Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L,
   Singh PB, Chang X, Schubert I, Jenewein T, Jacobsen SE:
   Dimethylation of histone H3 lysine 9 is a critical mark for DNA
   methylation and gene silencing in Arabidopsis thaliana.

51. Malagnac F, Barle E, Bender J: An Arabidopsis SET domain
   protein required for maintenance but not establishment of

52. Ebbs ML, Barle E, Bender J: H3 lysine 9 methylation is
   maintained on a transcribed inverted repeat by combined
   action of SUVH6 and SUVH4 methylestersases. Mol Cell

53. Ebbs ML, Bender J: Locus-specific control of DNA methylation
   by the Arabidopsis SUVH5 histone methylesterase. Plant

54. Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L,
   Schubert D, Patnaik D, Pradhan S, Goodrich J, Schubert I et al.: Dual
   histone H3 methylation marks at lysines 9 and 27 required
   for interaction with CHROMOMETHYLASE3. EMBO J 2004,
   23:4286–4296.

55. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J,
   Jacobsen SE: The SRA methyl-cytosine-binding domain links

56. Rountree MR, Selker EU: DNA methylation and the formation of

57. Zemach A, Zilberman D: Evolution of eukaryotic DNA
   methylation and the pursuit of safer sex. Curr Biol 2010,
   20:R780–785.
Review on the evolution of DNA methylation, methylation in different sequence contexts, gene body methylation, and transposon methylation in plants, animals, and fungi.

58. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild
   C, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: Shotgun
   bisulphite sequencing of the Arabidopsis genome reveals DNA
See annotation to [59**].

59. Lister R, O’Malley RC, Tonii-Filippini J, Gregory BD, Berry CC,
   Millar AH, Ecker JR: Highly integrated single-base resolution
This paper, together with [59**], presented the first whole-genome single-nucleotide resolution bisulfite sequencing studies for any organism. BS-Seq technology was used to detect methylation in regions of the genome (e.g., DNA genes and telomere) that are inaccessible to previous methods, and to study distribution, sequence context, and periodicity of methylation. These two papers also analyzed various mutants that have altered DNA methylation status and found that MET1 is mainly responsible
for the gene body CG methylation.

60. Zhang Xia, Yagai J, Sundaesran A, Cokus S, Chan SW, Chen H,
   Henderson IR, Shinn P, Pellegrini M, Jacobsen SE et al.: Genome-
   wide high-resolution mapping and functional analysis of DNA

61. Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE,
   Henikoff S: DNA methylation profiling identifies CG
   methylation clusters in Arabidopsis genes. Curr Biol 2005,
   15:154–159.

62. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S:
   Genome-wide analysis of Arabidopsis thaliana DNA
   methylation uncovers an interdependence between

63. Hisie TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L,
   Fischer RL, Zilberman D: Genome-wide demethylation of

64. Yoder JA, Walsh CP, Bestor TH: Cytosine methylation and the
   ecology of intragenomic parasites. Trends Genet 1997,

65. Chadavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H,
   between nucleosome positioning and DNA methylation.
This paper mapped nucleosome starting positions using high-throughput sequencing and showed that Arabidopsis exons are highly enriched for nucleosomes whereas introns are not. More interestingly, DNA methylases are elevated over nucleosomes, where all three types of methylation (CG, CHG, and CHH) exhibit a pronounced ten-base periodicity.

66. Laurent L, Wong E, Li G, Huynh T, Tsingos A, Ong CT, Low HM,
   Kim Sung KW, Rigoutsos I, Loring J et al.: Dynamic changes in
   the human methylome during differentiation. Genome Res 2010,
See annotation to [74**].


68. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-
   Bromage H, Tempst P, Lin SP, Allis CD et al.: DNMT3L connects
   unmethylated lysine 4 of histone H3 to de novo methylation of

69. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totsuki Y,
   Toyoda A, Ikawa M, Asada N, Koijima K, Yamaguchi Y, Ijiri TW
   et al.: DNA methylation of retrotransposon genes is regulated
   by Piwi family members MILI and MIW2 in murine fetal testes.

70. Riggs AD: X inactivation, differentiation, and DNA methylation.

71. Woo HR, Pontes O, Pikaard CS, Richards EJ: VIM1, a
   methylcytosine-binding domain protein required for centromeric

   UHRF1 plays a role in maintaining DNA methylation in

73. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA,
   Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K et al.: The SRA
   protein Np95 mediates epigenetic inheritance by recruiting

74. Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonii-
   Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al.: Human DNA
   methylestersomes at base resolution show widespread epigenomic
This paper and [66**] were the first two whole-genome single-nucleotide resolution bisulfite sequencing papers for human cells. They showed high levels of non-CG methylation in human embryonic stem (ES) cells. This paper identified differentially methylated regions (DMRs) across the genome between ES cells and fibroblasts. DNA methylation data were also correlated with gene expression, small RNA, and ChiP-Seq data.

75. Mosammaparast N, Shi Y: Reversal of histone methylation:
   • biochemical and molecular mechanisms of histone
Comprehensive review covering the current literature of histone demethylation studies. The demethylases are discussed in terms of their biochemistry, structure, enzymology, as well as their biological functions.