

PROBLEMS & PARADIGMS

Prospects & Overviews

Shaping eukaryotic epigenetic systems by horizontal gene transfer

Irina R. Arkhipova  | Irina A. Yushenova  | Fernando Rodriguez 

Marine Biological Laboratory, Josephine Bay
Paul Center for Comparative Molecular
Biology and Evolution, Woods Hole,
Massachusetts, USA

Correspondence

Irina R. Arkhipova and Fernando Rodriguez,
Marine Biological Laboratory, Woods Hole,
MA 02543, USA.
Email: iarkhipova@mbl.edu and
frodriguez@mbl.edu

Funding information

U.S. National Institutes of Health,
Grant/Award Number: R01GM111917; U.S.
National Science Foundation, Grant/Award
Number: MCB-2139001

Abstract

DNA methylation constitutes one of the pillars of epigenetics, relying on covalent bonds for addition and/or removal of chemically distinct marks within the major groove of the double helix. DNA methyltransferases, enzymes which introduce methyl marks, initially evolved in prokaryotes as components of restriction-modification systems protecting host genomes from bacteriophages and other invading foreign DNA. In early eukaryotic evolution, DNA methyltransferases were horizontally transferred from bacteria into eukaryotes several times and independently co-opted into epigenetic regulatory systems, primarily via establishing connections with the chromatin environment. While C5-methylcytosine is the cornerstone of plant and animal epigenetics and has been investigated in much detail, the epigenetic role of other methylated bases is less clear. The recent addition of N4-methylcytosine of bacterial origin as a metazoan DNA modification highlights the prerequisites for foreign gene co-option into the host regulatory networks, and challenges the existing paradigms concerning the origin and evolution of eukaryotic regulatory systems.

KEYWORDS

amino-methyltransferase, DNA methylation, epigenetic silencing, lateral gene transfer, N4-methylcytosine, regulatory evolution, transposable elements

INTRODUCTION

One of the most quintessential components of epigenetics is DNA methylation, that is, enzymatic addition of a methyl group to a nucleobase, which introduces subtle but chemically distinct changes into cytosines and adenines without changing the underlying genetic code. These changes are central to numerous biological processes that take advantage of the presence of a heritable and potentially removable mark on DNA, forming the basis for many epigenetic phenomena. Methylation is the most common form of DNA modification, and three main types of methylated bases have been discovered, with varying distribution patterns across prokaryotes and eukaryotes: N⁶-methyladenine, 5-methylcytosine, and N⁴-methylcytosine^[1–5] (Box 1; Figure 1A). Although methylated bases in RNA (N⁶-methyladenosine, N¹-methyladenosine, N⁶,2'-O-dimethyladenosine, 5-methylcytidine, 5-hydroxymethylcytidine) gained much prominence in recent years,

we do not consider epitranscriptomics here due to substantially different enzymology, substrate requirements, and biological effects.^[6]

In prokaryotes, the presence of a DNA modification is usually a hallmark of a host restriction-modification (R-M) system, which protects methylated bacterial DNA from phage infections by destroying unmethylated non-self DNA with a cognate restriction endonuclease (REase).^[16,17] R-M systems can travel horizontally across prokaryotic species as selfish genetic elements, with the same target sequence recognized by the REase and the MTase to safeguard genome integrity.^[18] Additionally, stand-alone or “orphan” MTases, such as *Dam*-like and *CcrM*-like, which are not associated with any REase, can be involved in DNA replication, mismatch repair, control of cell cycle, gene expression, and transposition.^[19–21] In eukaryotes, C5-methylation is overwhelmingly prevailing, while amino-methylation is far from typical. Equally atypical for eukaryotes are the corresponding N4/N6 modifying enzymes, homologues to which were sporadically reported in

Box 1. Methylated bases in DNA

Together with histone modification, chromatin remodeling, and small RNAs, DNA modification constitutes one of the cornerstones of epigenetics. In eukaryotes, the predominant form of DNA base methylation is C5-methylcytosine (5mC, m⁵dC), which represents the primary epigenetic modification involved in the control of crucial biological processes such as repression of transposable elements (TEs) and certain genes, genomic imprinting, X-chromosome inactivation, and early embryonic development.^[7,8] For a long time, investigations of epigenetic functions of methylated bases in eukaryotes were focused on 5mC, aptly named “the fifth base” of eukaryotic DNA due to its prevalence in plants and animals, especially in mammals. Besides the C5 position in the pyrimidine ring, the methyl group can also be added to the exocyclic amino group of adenine or cytosine (Figure 1, top). N6-methyladenine (6mA, m⁶dA, N⁶mA) represents the most abundant DNA modification in bacteria^[9,10] and was also reported in selected eukaryotes.^[11–13] In contrast, N4-methylcytosine (4mC, m⁴dC, N⁴mC) is relatively rare and was previously thought to occur exclusively in bacteria.^[11,14,15] Note that the corresponding nucleoside modifications in RNA are typically denoted as m⁶A and m⁴C, to facilitate discrimination from DNA. Superscripts correspond to the numbering of atoms; however, the use of superscripts is optional. No DNA MTases have yet been reported to act on RNA, and vice versa.

ciliates, fungi, and even a few mammals, although their functionality in the latter is questionable, as they have not been linked with N4/N6 base modifications in DNA.^[11,22,23]

While 5mC constitutes a well-recognized cornerstone of mammalian and plant epigenetics, it is more difficult to assign straightforward epigenetic roles to eukaryotic 6mA methylation marks, and the function, abundance and even presence of 6mA in DNA of multicellular eukaryotes have been debated on multiple occasions (see^[24,25] for review). Recently, the repertoire of methylated bases in eukaryotes was expanded to include the third modification, 4mC, added by N4-MTase which was horizontally transferred from a bacterial donor to the common ancestor of bdelloid rotifers about 60 million years ago.^[26] Its relatively recent recruitment into the host epigenetic system provides an opportunity to infer how the newly acquired DNA modifications can turn into epigenetic marks and to understand the requirements for a successful evolutionary transition from decorating the DNA double helix with a previously unseen chemical modification to its integration into the complex regulatory mechanics of the host cell. Since 4mC was previously known only in bacterial DNA, below we seek to explore how a eukaryotic genome can acquire a bacterial modification mark, how it can be reliably detected and experimentally validated, and how it

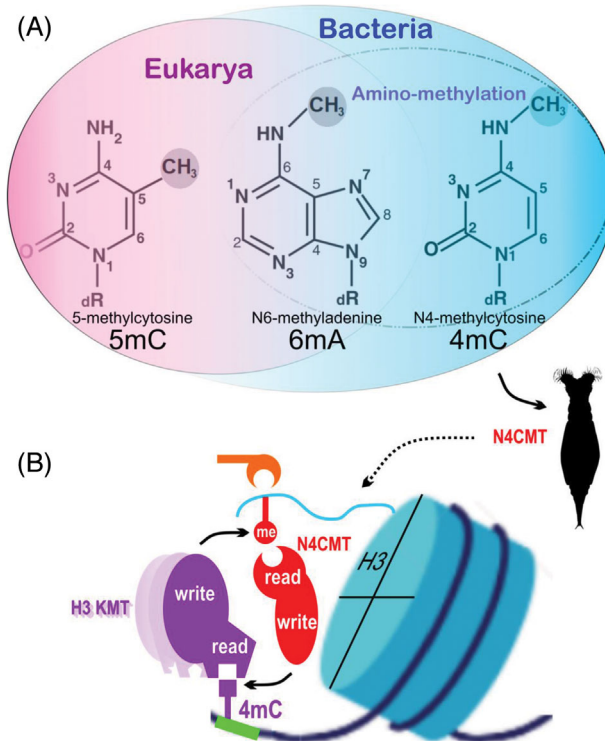


FIGURE 1 DNA methyl marks from bacteria to eukaryotes. (A) Three major types of DNA methylation. The atoms in the purine and pyrimidine rings are numbered. The methyl groups added to cytosine and adenine are circled. dR, deoxyribose. Coloring denotes the prevalence of modifications in different domains of life, with blue for bacteria, pink for Eukarya, and the overlap between them as blended color. The bottom panel (B), adapted from^[26], exemplifies recruitment of a horizontally transferred bacterial N4C-MTase into a eukaryotic epigenetic silencing system based on histone modifications. Both DNA and histone MTases (H3KMT) are composed of catalytic (“write”) and recognition (“read”) domains, which exist in several variants in H3KMT paralogs.

can become incorporated into the complex regulatory networks of a eukaryotic host to yield evolutionary innovations that did not exist in bacteria.

HORIZONTAL GENE TRANSFER FROM BACTERIA AS THE SOURCE OF EUKARYOTIC DNA METHYLTRANSFERASES

All known eukaryotic DNA MTases are of bacterial origin, having been acquired independently from different bacterial lineages and preserved by evolution following successful adaptation to their new eukaryotic host. In other words, they originated by horizontal gene transfer (HGT), which is also called lateral gene transfer (LGT); these terms are fully interchangeable and denote the transmission of genetic information by means other than parent-to-offspring (vertical). The evolutionary history of eukaryotic DNA MTases has been described in several reviews, which covered the origin of C5-MTases such as

Dnmt1 and Dnmt3 from bacterial Dcm enzymes,^[27] as well as phylogenetic origins of N⁶A-MTases, which were divided into three large groups (Group-1, M.MbolIA/M.MunI-like; Group-2, M.EcoKI/M.TaqI-like; Group-3, Dam-like).^[28]

In eukaryotes, HGT is a rare phenomenon in contrast to bacteria and has long been thought to be a relatively minor force in adaptive evolution in comparison with gene duplication, regulatory evolution, and so forth. However, certain metazoan taxa were discovered to be capable of capturing massive amounts of foreign genes, with bdelloid rotifers representing a prime example where 8%–10% of genes have originated from non-metazoan sources.^[29,30] Gradually, evidence began to accumulate that non-vertical processes such as HGT can add substantially to the adaptive potential of recipient taxa by bringing in previously absent functions.^[31] Nevertheless, most of these functions are confined to the so-called operational genes responsible for a specific biochemical reaction, not involving informational or other complex interconnected regulatory systems ("the complexity hypothesis").^[32,33]

Recently, two eukaryotic N⁴C-MTases were shown to be similarly acquired by HGT, as evident from complete lack of detectable eukaryotic homologs. In the phylum Rotifera, the enzyme N4CMT, fused to a eukaryotic chromodomain, helps to epigenetically suppress transposable elements (TEs) in rotifers of the class Bdelloidea^[26]; and an independent N4-MTase acquisition from bacteria occurred in the liverwort *Marchantia* (Viridiplantae; Streptophyta), where MpDN4MT1 is essential for spermiogenesis.^[34] In the latter case, however, no apparent cross-talk with histone-based epigenetics has been established, as the liverwort enzyme contains no extra domains and deposits 4mC marks rather indiscriminately, except when hampered by pre-existing 5mC marks.

To find out if other metazoans may have captured intact N4/N6-MTases from bacteria, we turned to the sensitive hidden Markov model (HMM) profile searches, which could reveal additional HGT instances. In the Pfam database,^[35] use of the PF01555 profile as a query returned 86 sequences from 33 eukaryotic species (Figure 2B, compared with nearly one-half of C5-MTases coming from eukaryotes). However, ORFs in 14 of these species lack catalytic residues, and nine more species were filtered out as bacterial contaminants (Table S1). The remaining 10 species were placed in the following taxonomic groups: one in Viridiplantae (green plants), three in Rhodophyta (red algae), one in Parabasalia (flagellated protists), three in Haptophyta (photosynthetic microalgae), one in Oomycetes (stramenopiles), and one in Glomeromycete fungi (Table 1). Upon clustering, however, only MTases from *Marchantiopsida* (Viridiplantae) showed similarity to rotifers, grouping with bacterial enzymes which methylate CpG targets and contain the SPPY catalytic motif characteristic of N4-MTases (Figure 2A,C). Other HGT candidates are clustered with N6-MTases retrieved from REBASE,^[36] and harbor the DPPY catalytic motif likely to methylate adenine (Figure 2C). The search is by no means exhaustive, as only annotated proteins in the Pfam database were inspected. Recent horizontal transfers can escape annotation due to poor recognition of alien ORFs by annotation pipelines trained on eukaryotic proteomes, evading detection. Nevertheless, the rarity of N4-MTase

acquisition/retention in comparison with N6-MTases is notable, and may reflect the known bias in amino-MTase distribution in the donor prokaryotic taxa and/or lack of successful functional recruitment.

Of note, 4mC represents a prominent modification in the genomes of giant DNA viruses, such as mollivirus and pandoraviruses, which infect amoebas of the genus *Acanthamoeba*.^[37] The corresponding N4-MTases (the same permuted Type II subtype β ^[38]) are scattered among bacterial type II N4-MTase clusters denoted by specific target sequences (Figure 2C), but are not paired with cognate REases, and are thought to be used for protection against R-M systems of co-infecting bacteria or viruses.

Out of the previously described groups of eukaryotic amino-MTases not presented here, most enzymes correspond to adenine N6-MTases.^[27,28] The only exception are the non-permuted MTases forming Clade 5 of Iyer et al.,^[28] found in Chlorophyta (green algae), SAR and Haptista. Most of these Clade 5 ORFs feature SPPY as the catalytic motif, making them good candidates for the role of cytosine N4-MTases (except for SCPY motif in unicellular green algae *Ostreococcus* spp. and GPPY in the dinoflagellate *Symbiodinium minutum*). Again, the non-permuted N4-MTases from giant viruses infecting amoebae (pandoraviruses) display homology to bacterial MTases with the same recognition sequence, rather than to eukaryotic hosts.^[37]

Note that it is not possible to reliably discriminate between the ability of an MTase to modify N4-cytosine or N6-adenine based on sequence homology alone, as they could be quite similar to each other in amino acid sequence and even functionally interchangeable.^[39] Indeed, 6mA and 4mC bases are chemically similar, while being quite distinct from 5mC (Figure 1A). Nonetheless, the homology-based approach can be used for preliminary assessment of N4 versus N6 enzyme specificity until further experimental validation (see below). So far, none of the SPPY-carrying MTases were shown to methylate N6-adenine, confirming that this signature motif can serve as a good indicator of specificity for N4-cytosine (the reverse does not apply, as there are many examples to the contrary). Recognition of any adjacent bases in the DNA target, however, can be inferred only tentatively, on the assumption that base specificity depends on amino acid residues in the target recognition domain (TRD) interacting with target bases, so that similarities in TRD should at least partially reflect similarities in the target sequence (e.g., sharing the CpG target, Figure 2C), but would have no correlation with bacterial phylogeny.

METHODOLOGICAL CHALLENGES IN IDENTIFYING NON-CANONICAL BASE MODIFICATIONS

The number of methods developed to detect canonical DNA methylation, either genome-wide or at single-gene level, has steadily increased over the past few decades. Most of these methods were focused on identifying methylation in eukaryotes, where 5mC DNA methylation is prevalent.^[40] The main approaches include digestion by REases differing by sensitivity to methylation; liquid chromatography; affinity enrichment of methylated DNA fragments using methyl-binding proteins or antibodies; and chemical conversion (e.g., with bisulfite)

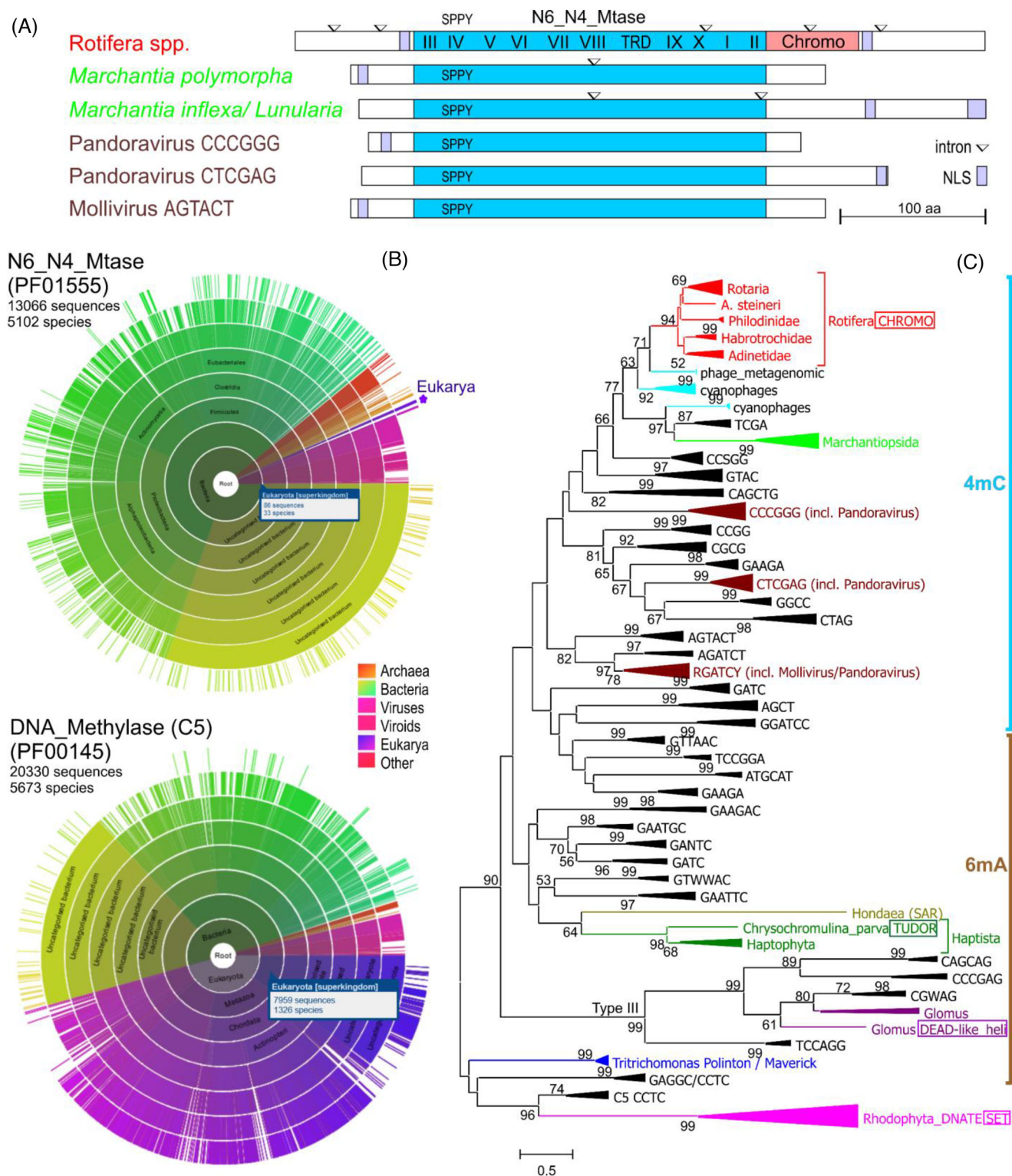


FIGURE 2 Domain structure and distribution of N6_N4_MTases. (A) Structural organization of eukaryotic N4C-MTases from rotifers, liverworts and giant viruses. Shown are conserved motifs I-X in the permuted MTase domain, intron positions (triangles), and nuclear localization signals (NLS) predicted via HMM.^[128] (B) Visualization of PF01555 distribution across species as a “sunburst” diagram from PFAM, with each ring (from inside to outside) showing the number of taxa within superkingdom, kingdom, phylum, class, order, family, genus, and species, respectively. The eukaryotic sector is marked with a purple asterisk. For comparison, the analogous distribution of PFAM family members for PF00145 (C5 DNA_methylase) is shown below. HMM search details are outlined in Table S1. (C) Clustering of eukaryotic MTases (in color) with their prokaryotic counterparts (in black) by the neighbor-joining algorithm. Prokaryotic MTases with their recognition sequences are from REBASE. N4C-MTases from bdelloid rotifer families, red; from liverworts, lime green; giant virus N4-MTases (in brown) are nested within the respective bacterial groups; cyanophages are highlighted in cyan. N6A-MTases from SAR, Haptista and Fungi are in olive, green, and purple, respectively. Putative parabasalid

TABLE 1 Eukaryotic species containing the N6_N4_MTase domain with intact catalytic residues.

Species	Taxonomy	Catalytic motif	Introns	NLS	Fused domain
<i>Adineta vaga</i>	Rotifera; Bdelloidea; Adinetidae	SPPY	+	+	Chromo
<i>Adineta ricciae</i>	Rotifera; Bdelloidea; Adinetidae	SPPY	+	+	Chromo
<i>Adineta steineri</i>	Rotifera; Bdelloidea; Adinetidae	SPPY	+		Chromo
<i>Macrotrachela quadricornifera</i>	Rotifera; Bdelloidea; Philodinidae	SPPY	+	+	Chromo
<i>Rotaria macrura</i>	Rotifera; Bdelloidea; Rotaria	SPPY	+	+	Chromo
<i>Rotaria magnacalcarata</i>	Rotifera; Bdelloidea; Rotaria	SPPY	+		Chromo
<i>Rotaria sordida</i>	Rotifera; Bdelloidea; Rotaria	SPPY	+	+	Chromo
<i>Rotaria sp. Silwood-1</i>	Rotifera; Bdelloidea; Rotaria	SPPY	+	+	Chromo
<i>Rotaria sp. Silwood-2</i>	Rotifera; Bdelloidea; Rotaria	SPPY	+		Chromo
<i>Habrotracha rosa</i>	Rotifera; Bdelloidea; Habrotrachidae	SPPY	+	+	Chromo
<i>Marchantia polymorpha</i>	Streptophyta; Marchantiopsida; Marchantiales	SPPY	+	+	nd
<i>Marchantia paleacea</i>	Streptophyta; Marchantiopsida; Marchantiales	SPPY	+	+	nd
<i>Marchantia inflecta</i>	Streptophyta; Marchantiopsida; Marchantiales	SPPY	+	+	nd
<i>Lunularia cruciata</i>	Streptophyta; Marchantiopsida; Lunulariales	SPPY	+		nd
<i>Hondaea fermentalgiana</i>	Sar; Stramenopiles; Bigyra; Thraustochytrida	DAPY	+		TPR
<i>Chrysochromulina tobinii</i>	Haptista; Haptophyta; Prymnesiales	DPPY	+	+	nd
<i>Chrysochromulina parva</i>	Haptista; Haptophyta; Prymnesiales	DPPY	–		Tudor
<i>Diacronema lutheri</i>	Haptista; Haptophyta; Pavloales	DPPY	–	+	AP2
<i>Emiliana huxleyi</i>	Haptista; Haptophyta; Isochrysidales	DPPY	+		nd
<i>Glomus cerebriforme</i>	Fungi; Mucoromycota; Glomeromycetes	DPPY	+		DEAD
<i>Trichomonas foetus</i> (TE)	Metamonada; Parabasalia; Trichomonadida	DPPY	–		nd
<i>Porphyra umbilicalis</i> (TE)	Rhodophyta; Bangiophyceae; Bangiales	DPPY	–	+	SET
<i>Chondrus crispus</i> (TE)	Rhodophyta; Florideophyceae; Gigartinales	DPPF/Y	–		SET
<i>Gracilariopsis chorda</i> (TE)	Rhodophyta; Florideophyceae; Gracilariales	DPPY	–	+	SET

Note: Shown are the species analyzed in Figure 2B. For complete information, see Table S1.
Abbreviation: nd, not detected by HMM.

(Box 2). Third-generation DNA sequencing technologies such as PacBio and Oxford Nanopore, in combination with improved bioinformatic approaches, have greatly increased the sensitivity and resolution of methylation analysis, and stimulated identification of low-abundance DNA methyl groups.^[41–43] These include the still controversial low-abundance 5mC in *Drosophila*,^[44,45] as well as 6mA (reviewed in^[24,25]) and the recently added 4mC marks^[26,34] in other eukaryotes, which comprise non-canonical base modifications, as opposed to the “canonical” abundant 5mC typical of mammalian genomes. Such non-canonical modifications often exhibit non-uniform spatial or temporal distribution (e.g., TE-associated or cell type-/developmental stage-restricted) and constitute a minor fraction of modified versus unmodified bases in the genome. While modern techniques offer new advantages for

detecting the elusive methylated bases, it is important to highlight potential pitfalls and possible ways to avoid artifacts in the analysis (Box 2).

Independently of the technique chosen for the detection of non-canonical modifications, contamination is a major problem when methylated DNA is of very low abundance in the target organism. Misincorporation of salvaged pre-methylated RNA or DNA nucleosides, either from m⁶A-rich cellular RNA or from DNA of commensal bacteria, by DNA polymerases has sometimes been invoked as a potential source of 6mA in eukaryotic DNA.^[25] By analogy, m⁴C could be misincorporated from mitochondrial rRNA, to which it is added by METTL15,^[62] or perhaps 4mC could be borrowed from rare bacterial DNA. Such misincorporation, however, is expected to be distributed

N6A-MTases (blue) are located within Polinton/Maverick transposons; SET-domain-containing MTases from red algae (magenta), with the closest match to C5-MTases, also belong to multicopy DNA transposons (Plavaka). Fusions of chromatin-associated domains to MTases are boxed. All MTases are classified by REBASE as permuted type II, except for the *Glomus*-containing group, which belongs to type III. Scale bar, amino acid substitutions per site. The tree format does not imply phylogenetic relationships between taxa. The uncollapsed version with the complete set of taxa is shown in Figure S1, and accession numbers are in Table S1.

Box 2. Methodological advantages and drawbacks in detecting non-canonical DNA base modifications

Several methods for 5mC detection have been successfully adapted for detection of other methylated bases, and novel approaches based on long-read technologies can be used to detect a broad range of modifications. However, it is important to highlight potential pitfalls associated with each of these methods, and the need to apply orthogonal methods.

DNA immunoprecipitation sequencing (DIP-seq or MeDIP-seq) is a common method for methylation detection and genome-wide mapping. Modern high-throughput sequencing techniques provide sufficient depth, identifying most of immunoprecipitated methylated DNA at relatively low cost. However, the specificity is limited by fragment size and by potential cross-reactivity of antibodies raised against specific methylated bases (anti-5mC, anti-6mA, anti-4mC) with other similar methylated bases, such as m⁶A or m¹A modifications in RNA, or with certain simple unmethylated motifs or secondary structures promoting off-target binding.^[46,47] Considering the limitations of antibody conjugation, one can address these problems by including input and IgG controls, as well as by using orthogonal non-antibody-based techniques. Bioinformatic curation in assembled reference genomes should be performed routinely to remove contaminants (e.g., DNA of commensal bacteria).^[48,49]

Chromatography-based techniques, such as LC-MS/MS (Liquid Chromatography-Tandem Mass Spectrometry), HPLC- (High-Performance-) and UHPLC-MS/MS (Ultra-High-Performance Liquid Chromatography-Tandem Mass Spectrometry), are highly specific and sensitive methods for bulk detection of non-canonical base modifications. Unlike DIP-seq, HPLC-MS/MS can precisely quantify each base signal (methylated/unmethylated ratio), and distinguish from methylated RNA products.^[14] Yet, HPLC-MS/MS results can be misinterpreted due to contaminated samples. Therefore, mass-spectrometry is a highly useful method that must be combined with other, sequence-based, methods when searching for rare base modifications. Commercial enzymes used to digest DNA may also carry contaminating bases, necessitating background measurements.^[50]

Bisulfite sequencing (BS-seq) provides a quantitative assay for methylated cytosine detection at base-resolution.^[51] Known as the gold standard for characterizing 5mC in eukaryotic genomes, it is based on deamination of unmethylated cytosines to uracil, leaving methylated cytosines intact. After treatment, DNA is PCR-amplified with a uracil-tolerant polymerase, causing uracil to convert to thymine. Sequences can be read directly by traditional Sanger or Illumina short-read sequencing through comparison to a reference or untreated sequence. However, bisulfite treatment can damage DNA, leading to degradation and PCR problems. Treating genomic DNA with TET (ten-eleven translocation) enzymes before bisulfite treatment makes it possible to characterize both 5mC and 4mC in bacterial genomes, although the differences are not always easily distinguishable.^[52]

PacBio SMRT and **Oxford Nanopore** single-molecule sequencing technologies can detect DNA methylation at base resolution, relying on changes in fluorescent pulse intervals or in electrolytic current signals, respectively.^[41–43] In contrast to BS-seq, single-molecule methods do not require base conversion to detect modifications. In SMRT-seq, base addition kinetics is measured during polymerization, and IPD (interpulse duration) ratio-based methylation profiling detects 6mA, 4mC, or other modifications such as 5-hydroxymethylcytosine (5-hmC) by measuring changes in the rate of polymerase processivity caused by modified bases. Initially applied to bacteria,^[10] the SMRT technology also boosted detection of non-canonical base modifications in eukaryotes,^[24,26,34] although it cannot distinguish between 6mA and 1mA, or reliably call modifications near cytosines.^[46,53,54]

The biggest SMRT-seq drawback is a high false discovery rate (FDR) at low modified/unmodified base ratios.^[46,50] At borderline values (under 1%), deeper coverage is needed to reduce the FDR: as read coverage increases, the variance of normal distributions of IPDs at the single-molecule level decreases, providing increased power for separation between methylated and non-methylated bases.^[55] Contaminating DNA in reference genomes was shown to inflate 6mA/A ratios in fruit flies, plants and humans, and a recent quantitative metagenomic approach applies machine learning to minimize artifacts.^[56,57] The most reliable way to reduce FDR is to use unmethylated DNA (via demethylation or PCR amplification) as a control.^[58,59] Although far from cost-effective, this approach demonstrated that multiple IPD shifts in *C. elegans* were caused by organism-specific idiosyncratic sequence contexts,^[60] in addition to previously known obstacles such as non-B DNA motifs.^[61]

randomly across the genome, without regard to annotated genomic features.

Improvement of methods for 4mC detection in eukaryotes is only beginning. Its partial sensitivity to bisulfite deamination makes discrimination from 5mC possible but cumbersome.^[34,63] While nanopore sequencing and modified bisulfite sequencing (4mC-TAB-seq) were used to detect 4mC in bacteria,^[52,64] a promising approach is fur-

ther development of chemical conversion via nitrosylation, followed by PCR amplification.^[65] However, application of this method to eukaryotes may be complicated by DNA degradation and by PCR amplification biases. Resistance of 4mC to enzymatic deamination by APOBEC3A (4mC-AMD-seq) was used to discriminate 4mC from C and 5mC in the radioresistant bacterium *Deinococcus radiodurans*.^[66] This method avoids excessive DNA degradation and can be applied to

low-input samples. Further improvement of 4mC detection methods in eukaryotes will be essential for establishment of validated 4mC reference datasets for training of bioinformatic pipelines, and for eventual elucidation of its biological functions.

CHALLENGES IN ASSIGNMENT AND VALIDATION OF EUKARYOTIC METHYLTRANSFERASE FUNCTION

Since 2017, over 30 publications aimed at characterizing 4mC distribution in eukaryotic genomes were devoted to continuous improvement of 4mC prediction accuracy using machine learning approaches ([67,68] and references therein). However, the value of these methods is uncertain, as they were developed and trained on reference datasets^[69,70] which have not been rigorously proven to harbor the 4mC mark. As explained above, the principal techniques employed for detection of 4mC and database creation, such as SMRT-seq or UHPLC-MS/MS, are prone to inaccuracies (high signal-to-noise ratio, bacterial and/or RNA contamination).^[50] Thus, in addition to confirming the presence of 4mC modified bases in sequenced genomes by orthogonal methods, it is crucial to identify the corresponding MTase enzyme before making any claims regarding the presence of 4mC in a genome. If 4mC is identified by SMRT-seq analysis in a genome containing only C5-MTase homologs (which include DNMTs),^[71–73] the most natural explanation is the erroneous identification of 5mC marks as 4mC by the software used.

Amino-MTases responsible for incorporation of 4mC and 6mA into DNA are somewhat less conserved than C5-MTases, and are not easily distinguishable from each other in silico. Both C5- and amino-MTases contain the catalytic domain demarcated by 10 conserved motifs I-X and the variable DNA target recognition domain (TRD) (Figure 2A). The subdomains can be permuted in several MTase types,^[74,75] but it is difficult to predict methylation specificity based solely on the order and sequence of motifs. While DNA C5-MTases reliably group together, the diversity of amino-MTases also includes MTases acting on RNA and/or proteins (e.g., HemK family).^[28] Based on multiple sequence alignments and hidden Markov models (HMM) profile searches, most amino-MTases would be assigned to the N6_N4_MTase domain family (PF01555) originating from the Pfam protein database.^[35] The general mechanism of catalysis with base flipping, implied by the [DNSH]PP[YFW] amino acid string, is reviewed in Refs.^[28, 76, 77]

Although the current algorithms for genome assembly are generally accurate, cases of misassembly still occur, implying that MTase DNA sequences should be verified using additional genomic or transcriptomic data. Homology searches are used to identify novel MTase sequences following prediction of protein-coding genes in a de novo genome assembly (Figure 3). Positive MTase hits must be checked for the presence of eukaryotic genes in the immediate genomic environment to rule out bacterial origin, even for large contigs, as entire endosymbiont genomes can be assembled with sufficient coverage (Table S1). Typically, intron accumulation constitutes good evidence for domestication of a formerly bacterial gene. REBASE provides a source

of MTases with known recognition sequences and methylation specificity, which can be used to create a multiple sequence alignment.^[36] To correctly assign a novel MTase candidate to the appropriate group, it is imperative to identify catalytic amino acids, which are part of motif IV. Most N4C-MTases have the SPPY catalytic string. With sufficient representation of amino- and C5-MTases and simple clustering approaches, any candidate MTase may be tentatively assigned to 4mC-, 6mA- or C5-modifying types.

No matter how reliable a protein alignment is, it cannot serve as proof that a novel MTase possesses the predicted enzymatic activity. Only experimental verification can decisively assess the specificity of a eukaryotic MTase (Figure 3). To determine whether an MTase shows enzymatic activity, and if so, to characterize it, biochemical approaches should be employed. Purifying the putative MTase protein directly from host cells, or producing a recombinant enzyme in a heterologous expression system, are the easiest solutions. However, it is not always possible to obtain pure protein of sufficient quality and quantity for in vitro studies. If so, one can perform indirect studies by expressing putative MTases in bacteria or tissue culture and analyzing DNA methylation status. Substitution of catalytic residues on either side of the prolines in the [DNSH]PP[YFW] domain to alanines should eliminate the corresponding methyl marks from DNA in control assays.^[78]

In vivo knockout/knockdown (KO/KD) experiments became the de facto gold standard for validation of biological function and establishment of causal links between phenotypes and candidate loci. However, KO/KD of the corresponding gene in vivo is possible mostly in model organisms, and may yield inconclusive results due to indirect and off-target effects on other cellular components, and/or co-existence of several methyl marks. Indeed, N6AMT1 was implicated in N⁶A methylation based on KD experiments,^[79] but is a HemK-like methylase acting on N⁵ in glutamine.^[80–82] Similarly, the candidate N6AMT2 was shown in vitro to act not on N⁶A in DNA, but on N⁵ in lysine of EF1A.^[83] On the other hand, using recombinant enzymes paired with their catalytic mutants (change to APPA in motif IV) can often give a clear answer about MTase specificity and recognition sequence. Use of [³H]-AdoMet or validated anti-4mC/anti-6mA antibodies in combination with DNA substrates to detect methylation in vitro are feasible approaches, although optimization of experimental conditions may be required.^[26,84] In sum, although assignment of prokaryotic enzyme specificity is now largely achieved via high-throughput analysis of genomes and methylomes,^[85,86] proper assignment of specificity to novel eukaryotic MTases requires a thorough manual curation of database search outputs, as well as experimental verification of sequence specificity and DNA modification activity.

PREREQUISITES FOR FUNCTIONAL RECRUITMENT BY EUKARYOTIC HOSTS

If 4mC in eukaryotes is so rare that it can be regarded as more of an evolutionary oddity than as a recurring phenomenon, it may be asked what we can learn from it about the general process of horizontal

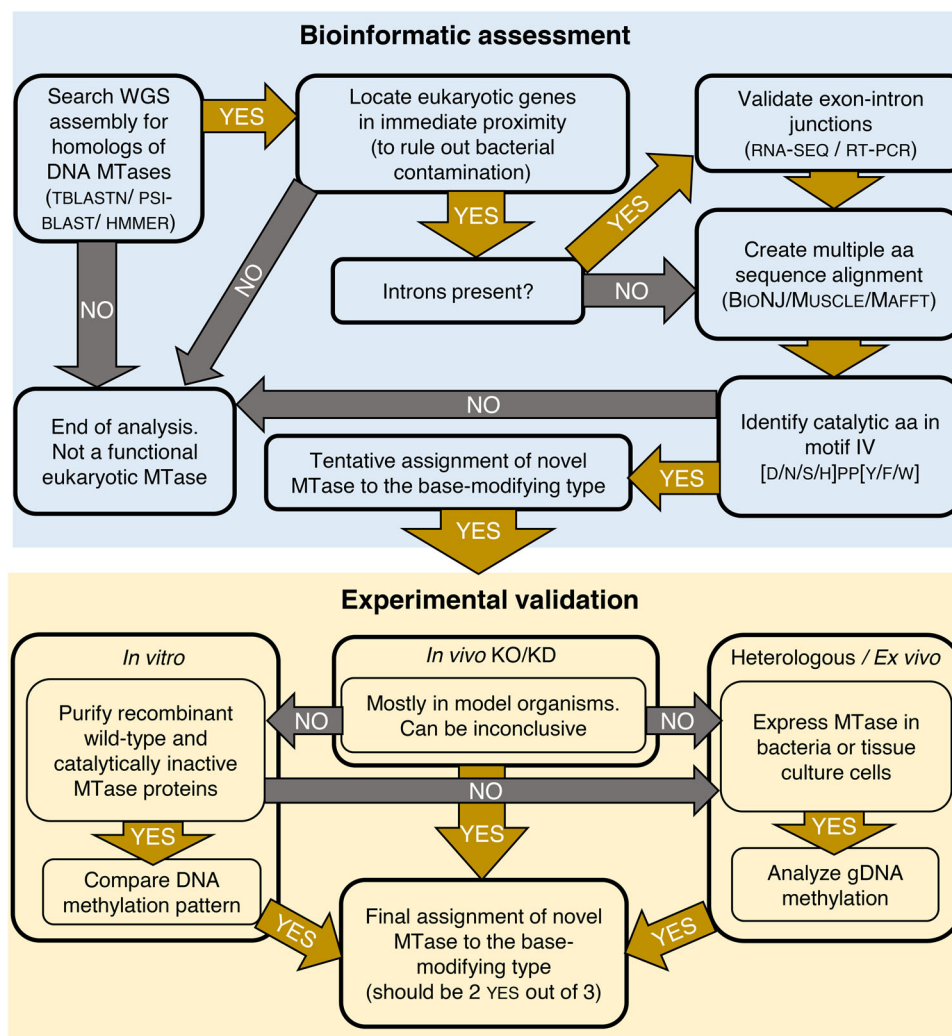


FIGURE 3 A flowchart outlining the required steps in identification and validation of novel DNA amino-MTases in eukaryotic genomes. WGS, whole-genome sequence; aa, amino acid. The [DNSH]PP[YFW] amino acid string in motif IV can serve as an initial indicator: most N4C-MTases have the SPPY catalytic string, while most N6A-MTases feature DPPY that distinguishes them from eukaryotic N6AMT1 (NPPY, N⁵-Gln), N6AMT2 (DPPY/F, N⁵-Lys), or METTL4-like enzymes (DPPW, also seen in METTL3/IME4-like m⁶A-RNA MTases).

acquisition of non-canonical DNA modifications and their subsequent engagement with pre-existing eukaryotic networks. When an MTase is transferred from bacteria, it would initially be intronless, but can be **expected to acquire introns** (Table 1; Figure 2A). Furthermore, to act on DNA, it should become capable of entering eukaryotic nuclei via **acquisition of a nuclear localization signal** (NLS). Indeed, NLS is found in most eukaryotic enzymes shown in Figure 2A, including giant viruses, in N- or C-terminal extensions to the MTase domain. Although an MTase could retain some target specificity inherited via its TRD, there is no a priori reason to expect that a bacterial enzyme would exhibit any regional specificity in a eukaryotic host without first acquiring the respective domain. Thus, the most natural adaptation that a bacterial DNA MTase could undergo is the transition from modifying nucleosome-free DNA to acting on DNA packaged into chromatin via **fusion to chromatin-binding domains**, signifying a substantial departure from bacterial circular chromosomes organized into a membrane-less nucleoid consisting mainly of DNA. Finally, if the methyl mark on a

stretch of DNA is expected to exert a biological function, it should be recognized by another host component, unless it affects the physical properties of the DNA itself. At this step, gene duplication can play an important role, allowing adaptive evolution of newly emerging paralogs into **specialized interacting partners**.

The above steps can be traced over evolutionary time at different timescales. In case of eukaryotic C5-MTases, functional recruitment occurred such a long time ago that the ancestry of the corresponding orthologs can be traced back to early eukaryotes, when the present-day domain configurations were established for Dnmt1- and Dnmt3-like MTases to ensure interaction with chromatin and accessory proteins.^[27] It is somewhat easier to observe the initial steps of MTase recruitment if the acquisition is more recent and can be measured in tens rather than hundreds of millions of years ago.

Enzymes acting on DNA in a eukaryotic cell acquire the means for interacting with DNA packaged into chromatin largely via domain fusion. Acquisition of the chromodomain (CHD) was the decisive

step for the bdelloid amino-MTase, enabling it to “read” silent chromatin marks located over TEs and tandem repeats, such as H3K9me3 and H3K27me3, and to deposit 4mC marks over these regions (Figure 1B).^[26] This is not the only example of a successful MTase-CHD fusion: in plants, CHD was inserted within a C5-MTase to form chromomethylases (CMT), enzymes present throughout the plant kingdom. Here, the degree of convergent evolution extends even further: while bdelloids contain a SETDB1 (SET domain bifurcated) histone H3 lysine (H3K) MTase variant with preference for “reading” 4mC marks (Figure 1B), the plant CMT3, which methylates C⁵ in the CHG context, forms a similar self-reinforcing feedback loop with the histone lysine MTase KRYPTONITE (KYP), which methylates H3K9 via its SET domain, allowing further H3K9me recognition and C5-methylation by CMT3.^[87]

While this feedback loop design is similar between plants and rotifers, there are also other differences besides C5 versus N4C methylation. KYP recognizes 5mC via its SAD/SRA DNA-binding domain, which performs 5mC base flipping for discrimination. In contrast, SETDB1 paralogs in rotifers contain the MBD/TAM DNA-binding domain, which is not capable of base flipping, but can interact directly with the major groove of DNA in the nucleosome context.^[88] Despite MBD being a generic DNA-binding domain, which does not always require methyl groups for DNA-protein interaction, development by bdelloids of even a modest preference for N⁴-methylated base by a paralog may have had a substantial cumulative effect. Similarly, the SRA domain of UHRF2 binds hemihydroxymethylated and fully hydroxymethylated DNA approximately 1.5 and 3.2 times more tightly than hemi-methylated DNA, respectively, with slight alterations in the binding pocket providing enough discriminatory power to classify it as 5hmC-specific reader in comparison with 5mC-specific UHRF1, the Dnmt1 cofactor.^[89]

As known from the highly complex domain architecture of the plant and mammalian *Dnmt1/Dnmt3* orthologs, a large variety of chromatin-interacting domains in addition to the chromodomain can be fused to MTases to ensure productive interactions with chromatin.^[87,90] In histone KMTases, a similarly high variety of domains can be observed, capable of interacting with repressive or active chromatin marks, and sometimes with both. For example, the Tudor domain can bind both H3K4me3 and H4K20me3 active marks,^[91,92] while within SETDB1 the triple Tudor domain binds to doubly modified H3 containing the repressive H3K9me3 and the active H3K14ac marks, forming bivalent chromatin.^[93] In bdelloid rotifers, although the interactions of the triple Tudor domain in SETDB1 paralogs have not yet been examined, the silent marks for H3K9me3 and H3K27me3 largely overlap, indicating the lack of a strict subdivision of heterochromatin into constitutive and facultative.^[26] While TE transcriptional activity shows a downward trend in H3Kme3-marked areas, it is possible that some of these regions may contain other marks indicative of bivalent nature.

An important factor contributing to epigenetic recruitment may be **DNA accessibility** provided by chromatin remodelers. In flowering plants and mammals, nucleosomes typically obstruct action of most types of DNA MTases, which preferentially act on nucleosome-free DNA in all sequence contexts, unless nucleosomes are de-compacted

by remodelers such as DDM1/Lsh.^[94] However, in deep-branching marine algae C5-methylation by other Dnmt-like MTases is mostly confined to the inter-nucleosomal linker region, likely contributing to nucleosome positioning.^[95] Linker preference is also observed for symmetric N⁶A-methylation in ciliates, which is performed by a member of the MT-A70 family of amino-MTases.^[58] Bdelloids do have a *Ddm1* ortholog and the corresponding capacity to remodel nucleosomes for efficient methylation, and amino-MTases depend on the same base-flipping mechanism for cytosine methylation as C5-MTases. On the other hand, recognition of DNA methyl groups in the major groove by an MBD/TAM domain in rotifer SETDB1-like H3K9-KMTases does not involve base flipping and could potentially be performed in the nucleosome context. In mammals, SETDB1 proteins are best studied in the context of KRAB-KAP1 and HUSH complexes which establish TE repression.^[96–98] In contrast, the SAD/SRA domain in H3K9-KMTases, such as KYP and SUVH9, and the Dnmt1 cofactor UHRF1, requires base flipping for recognition of C5-methylated bases in flowering plants and mammals, enhancing the need for chromatin remodeling to ensure H3K9 methylation.^[87]

It should be interesting to find out in what ways other domain combinations, which differ from those in the best-studied systems, can **organize the cross-talk between DNA and histone modifications**. Domain configuration of Dnmt1-like C5-MTases has remained relatively unchanged throughout the evolution of multicellular eukaryotes, while Dnmt3-like MTase domain architectures were configured independently in plants and animals for de novo C5 methylation.^[27,99,100] At the same time, other metazoan taxa with more taxonomically restricted and/or more evolutionarily recent MTases should offer additional opportunities to explore epigenetically relevant domain configurations. In addition to C5- and N6-MTase types identified in previous studies,^[27,28] Figure 2C outlines several previously unreported examples of taxa which may have evolved unconventional epigenetic systems upon fusion with chromatin-interacting domains: the C-terminal Tudor domain in a putative N⁶A-MTase from *Chrysochromulina parva*, a phytoplanktonic haptophyte, and the C-terminal SET domain in a clade of multicopy DNA transposons from red algae (*Chondrus crispus*, *Porphyra umbilicalis*, *Gracilariopsis chorda*). The latter clade may represent an interesting case of C5-specificity acquisition by amino-MTases, as the closest REBASE matches are C5-specific MTases recognizing CCTC, although the DPPY/F motif is typical of N6-MTases. With the development of more accurate methods permitting base-pair resolution of non-canonical DNA modifications,^[56,66] application of such methods to eukaryotic genomes with unconventional MTase types may yield interesting insights into the epigenetics of understudied eukaryotes.

POTENTIAL ROLES OF 4mC IN EUKARYOTIC GENOMES

Shortly after the discovery of 4mC in bacteria, it was noted that, in contrast to 5mC, this base modification is highly resistant to deamination.^[101,102] Cytosine deamination (C → U and especially

5mC → T) is highly mutagenic, and the prevalence of 4mC in thermophilic bacteria is widely believed to serve as protection from heat-induced deamination, although examples to the contrary are also known.^[103] Possibly, part of the reason why 5mC is not as prevalent in bacteria as it is in eukaryotes could be the higher degree of bacterial exposure to external environmental stresses and temperature shifts. Bdelloid rotifers are well-known for their extraordinary resistance to harsh environmental conditions, and are thriving in extreme environments from glaciers to acidic hot springs.^[104–107] Thus, the avoidance of cytosine deamination, when combined with benefits from epigenetic TE silencing, may be beneficial in terms of lower mutagenicity. In the radioresistant bacterium *D. radiodurans*, 4mC is thought to contribute to genome stability, as its removal yielded higher spontaneous mutation frequencies and enhanced recombination and transformation efficiencies.^[108] More esoteric effects of 4mC on DNA include preferential formation of pyrimidine(6-4)pyrimidone photoproducts resistant to hydrolysis, rather than cyclobutane pyrimidine dimers favored by 5mC, with the potential to considerably change DNA UV photochemistry.^[109,110]

Not too many studies have investigated the direct influence of DNA modifications on physical properties of DNA, although in early works researchers reported destabilization of DNA double helix by either 6mA or 4mC, slightly lowering the T_m ,^[111–113] while more recent studies reported that 5mC reduces DNA flexibility, disfavoring nucleosomes.^[114,115] Although these changes may affect nucleosome stability in vitro, the action of nucleosome remodelers may neutralize any such effects in vivo.^[58] Thus, it is likely that recruitment of a non-canonical base is best realized via its potential to serve as a covalently attached mark, which is recognizable as a chemically distinct DNA alteration by host proteins.

Such recognition mechanism, however, if not pre-existing in the host cell or brought in externally, must evolve within the new host, implying that a horizontally transferred MTase should survive through the period of adaptive evolutionary response from the host. For efficient use as an epigenetic mark, one of the prerequisites is the addition of a chromatin-interacting domain via fusion to a resident eukaryotic gene. Different paths to this structural reorganization could be entertained: direct fusion as DNA, for example, by introducing a double-strand break (DSB) with subsequent resection and non-homologous end-joining (NHEJ) to accomplish in-frame fusion, or via illegitimate recombination with unequal exchange; and/or through alternative splicing, with utilization of a cryptic splice site outside the MTase moiety and an existing splice site within a multi-domain protein, with subsequent re-insertion of a chimeric cDNA copy into the host genome. Bdelloids indeed undergo frequent DSB formation and NHEJ repair following repeated cycles of desiccation and rehydration, and harbor potentially active retrotransposable elements that could produce cDNA copies *in trans*.^[116–119]

Initial MTase survival, however, may depend on a possible function not requiring incorporation into the host epigenetic system. This has apparently happened in the liverwort *Marchantia polymorpha*, where N4-MTase has somehow acquired a function critical for sperm maturation via an unknown mechanism not dependent on domain fusion and

integration with the pre-existing 5mC methylation systems. In fact, the enzyme is uniformly labeling all sequences but TEs, which are already covered by dense 5mC marks, precluding 4mC addition to the same cytosine, which would be damaging to DNA.^[34,120] It remains to be established how recent the 4mC acquisition was in liverworts, although the shared intron was acquired prior to the split of the orders Marchantiales and Lunulariales (Figure 2A,C). It remains possible that N4CMT in bdelloid rotifers was initially retained for its intrinsic affinity for certain motifs, and was recruited for epigenetic TE suppression at a later stage when the fusion to CHD occurred.

CONCLUSIONS AND PROSPECTS

DNA methylation is a heritable alteration that can be superimposed onto the underlying genetic code without changing it. For a long time, studies of eukaryotic DNA methylation were focused on 5mC and the corresponding C5-MTases. After recent detection of 6mA in multiple eukaryotes, a flurry of publications started to investigate 6mA and potential N6-MTases. Here, we are adding the third methyl mark, 4mC, to the eukaryotic repertoire, and begin to explore 4mC/N4-MTase occurrence throughout eukaryotes and its potential to shape epigenetic landscapes and to participate in TE silencing. While TE suppression has always been an underlying theme in most epigenetic studies,^[121–123] a newly acquired MTase would not specifically target TEs without undergoing evolutionary adaptation.

HGT is mostly thought to involve the “operational” genes,^[32] and is not ordinarily expected to contribute to the complex regulatory systems of the host. Early horizontal recruitments of C5-MTases into the eukaryotic epigenetic system predated the divergence of the eukaryotic crown group into animal, fungal and plant kingdoms, but post-dated the histone modification system, which evolved in the last eukaryotic common ancestor.^[27,124] Almost invariably, MTase recruitment implies acquisition of additional protein domains required for interaction with nuclear DNA which is organized into chromatin. It is fascinating to observe how convergent evolution operated on relatively recent time scales when establishing connections of an introduced bacterial MTase with other epigenetic layers involved in TE silencing. Another prerequisite to this paradigm-shifting HGT role in re-shaping eukaryotic regulatory networks is the availability of gene duplications to facilitate evolution of regulatory loops, which can stimulate cross-talk in different directions to form multi-layered epigenetic systems.

Finally, it is worth re-emphasizing that future improvements of base modification detection techniques should rely on high-confidence training datasets that would not include data from species in which the presence of a given modification is in doubt. More exotic DNA modifications in addition to the bacterial “big three” are being reported, although the enzymatic basis for their deposition has not always been elucidated.^[125–127] Identification of the corresponding enzymes capable of adding such modifications should achieve highest priority in future studies exploring unconventional epigenetic systems across the eukaryotic tree of life.

ACKNOWLEDGMENTS

This work was supported by grants from the U.S. National Institutes of Health to I.A. (R01GM111917) and the U.S. National Science Foundation to I.A. and I.Y. (MCB-2139001).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study were obtained from the following public databases: Genbank, <https://www.ncbi.nlm.nih.gov/>; Pfam, at <https://pfam-legacy.xfam.org> until early 2023 and now integrated with InterPro <https://www.ebi.ac.uk/interpro/>; and REBASE, <https://rebase.neb.com/rebase/>; and are available as supplementary material to this article.

ORCID

Irina R. Arkhipova  <https://orcid.org/0000-0002-4805-1339>

Irina A. Yushenova  <https://orcid.org/0000-0001-6291-6215>

Fernando Rodriguez  <https://orcid.org/0000-0003-4044-8734>

REFERENCES

- Wyatt, G. R. (1950). Occurrence of 5-methylcytosine in nucleic acids. *Nature*, 166(4214), 237–238. <https://doi.org/10.1038/166237b0>
- Dunn, D. B., & Smith, J. D. (1958). The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochemical Journal*, 68(4), 627–636. <https://doi.org/10.1042/bj0680627>
- Vanyushin, B. F., Belozersky, A. N., Kokurina, N. A., & Kadirova, D. X. (1968). 5-methylcytosine and 6-methylamino-purine in bacterial DNA. *Nature*, 218(5146), 1066–1067. <https://doi.org/10.1038/2181066a0>
- Vanyushin, B. F., Tkacheva, S. G., & Belozersky, A. N. (1970). Rare bases in animal DNA. *Nature*, 225(5236), 948–949. <https://doi.org/10.1038/225948a0>
- Janulaitis, A., Klimašauskas, S., Petrušyte, M., & Butkus, V. (1983). Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. *FEBS Letters*, 161(1), 131–134.
- Li, X., Xiong, X., & Yi, C. (2017). Epitranscriptome sequencing technologies: decoding RNA modifications. *Nature Methods*, 14(1), 23–31. <https://doi.org/10.1038/nmeth.4110>
- Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, 31(2), 89–97. <https://doi.org/10.1016/j.tibs.2005.12.008>
- Edwards, J. R., Yarychivska, O., Boulard, M., & Bestor, T. H. (2017). DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin*, 10, 23. <https://doi.org/10.1186/s13072-017-0130-8>
- Varma, S. J., Calvani, E., Grüning, N.-M., Messner, C. B., Grayson, N., Capuano, F., Müllender, M., & Ralser, M. (2022). Global analysis of cytosine and adenine DNA modifications across the tree of life. *Elife*, 11, e81002. <https://doi.org/10.7554/eLife.81002>
- Blow, M. J., Clark, T. A., Daum, C. G., Deutschbauer, A. M., Fomenkov, A., Fries, R., Froula, J., Kang, D. D., Malmstrom, R. R., Morgan, R. D., Posfai, J., Singh, K., Visel, A., Wetmore, K., Zhao, Z., Rubin, E. M., Korlach, J., Pennacchio, L. A., & Roberts, R. J. (2016). The epigenomic landscape of prokaryotes. *PLOS Genetics*, 12(2), e1005854. <https://doi.org/10.1371/journal.pgen.1005854>
- Hattman, S. (2005). DNA-[adenine] methylation in lower eukaryotes. *Biochemistry (Moscow)*, 70(5), 550–558.
- Luo, G.-Z., Blanco, M. A., Greer, E. L., He, C., & Shi, Y. (2015). DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? *Nature Reviews Molecular Cell Biology*, 16(12), 705–710. <https://doi.org/10.1038/nrm4076>
- Heyn, H., & Esteller, M. (2015). An adenine code for DNA: A second life for N6-methyladenine. *Cell*, 161(4), 710–713. <https://doi.org/10.1016/j.cell.2015.04.021>
- Ratel, D., Ravanat, J.-L., Berger, F., & Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays*, 28(3), 309–315. <https://doi.org/10.1002/bies.20342>
- Wion, D., & Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature Reviews Microbiology*, 4(3), 183–192. <https://doi.org/10.1038/nrmicro1350>
- Meselson, M., & Yuan, R. (1968). DNA restriction enzyme from *E. coli*. *Nature*, 217(5134), 1110–1114.
- Linn, S., & Arber, W. (1968). Host specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage fd replicative form. *Proceedings of the National Academy of Sciences of the United States of America*, 59(4), 1300–1306. <https://doi.org/10.1073/pnas.59.4.1300>
- Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research*, 29(18), 3742–3756. <https://doi.org/10.1093/nar/29.18.3742>
- Boye, E., & Løbner-Olesen, A. (1990). The role of dam methyltransferase in the control of DNA replication in *E. coli*. *Cell*, 62(5), 981–989. [https://doi.org/10.1016/0092-8674\(90\)90272-g](https://doi.org/10.1016/0092-8674(90)90272-g)
- Roberts, D. (1985). IS10 transposition is regulated by DNA adenine methylation. *Cell*, 43(1), 117–130. [https://doi.org/10.1016/0092-8674\(85\)90017-0](https://doi.org/10.1016/0092-8674(85)90017-0)
- Wright, R., Stephens, C., & Shapiro, L. (1997). The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *Journal of Bacteriology*, 179(18), 5869–5877.
- Karrer, K. M. (2002). Methylation of adenine in the nuclear DNA of *Tetrahymena* is internucleosomal and independent of histone H1. *Nucleic Acids Research*, 30(6), 1364–1370.
- Ratel, D., Ravanat, J.-L., Charles, M.-P., Platet, N., Breuillaud, L., Lunardi, J., Berger, F., & Wion, D. (2006). Undetectable levels of N6-methyl adenine in mouse DNA: Cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. *FEBS Letters*, 580(13), 3179–3184. <https://doi.org/10.1016/j.febslet.2006.04.074>
- Luo, G.-Z., & He, C. (2017). DNA N(6)-methyladenine in metazoans: Functional epigenetic mark or bystander? *Nature Structural & Molecular Biology*, 24(6), 503–506. <https://doi.org/10.1038/nsmb.3412>
- Boulias, K., & Greer, E. L. (2022). Means, mechanisms and consequences of adenine methylation in DNA. *Nature Reviews Genetics*, 23(7), 411–428. <https://doi.org/10.1038/s41576-022-00456-x>
- Rodriguez, F., Yushenova, I. A., DiCorpo, D., & Arkhipova, I. R. (2022). Bacterial N4-methylcytosine as an epigenetic mark in eukaryotic DNA. *Nature Communications*, 13(1), 1072. <https://doi.org/10.1038/s41467-022-28471-w>
- Iyer, L. M., Abhiman, S., & Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. *Progress in Molecular Biology and Translational Science*, 101, 25–104. <https://doi.org/10.1016/B978-0-12-387685-0.00002-0>
- Iyer, L. M., Zhang, D., & Aravind, L. (2016). Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays*, 38(1), 27–40. <https://doi.org/10.1002/bies.201500104>
- Gladyshev, E. A., Meselson, M., & Arkhipova, I. R. (2008). Massive horizontal gene transfer in bdelloid rotifers. *Science*, 320(5880), 1210–1213.
- Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E. G. J., Hejnal, A., Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthélémy, R.-M., Bast, J., Bazykin, G. A., Chabrol, O., Couloux, A., Da Rocha,

- M., Da Silva, C., Gladyshev, E., ... Van Doninck, K. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, 500(7463), 453–457. <https://doi.org/10.1038/nature12326>
31. Gabaldón, T. (2020). Patterns and impacts of nonvertical evolution in eukaryotes: A paradigm shift. *Annals of the New York Academy of Sciences*, 1476(1), 78–92. <https://doi.org/10.1111/nyas.14471>
 32. Jain, R., Rivera, M. C., & Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences*, 96(7), 3801–3806. <https://doi.org/10.1073/pnas.96.7.3801>
 33. Cohen, O., Gophna, U., & Pupko, T. (2011). The complexity hypothesis revisited: Connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution*, 28(4), 1481–1489. <https://doi.org/10.1093/molbev/msq333>
 34. Walker, J., Zhang, J., Liu, Y., Vickers, M., Dolan, L., Nakajima, K., & Feng, X. (2021). Extensive N4 cytosine methylation is essential for *Marchantia* sperm function. *BioRxiv*, 2021.02.12.428880. <https://doi.org/10.1101/2021.02.12.428880>
 35. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419. <https://doi.org/10.1093/nar/gkaa913>
 36. Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, 43(Database issue), D298–D299. <https://doi.org/10.1093/nar/gku1046>
 37. Jeudy, S., Rigou, S., Alempic, J.-M., Claverie, J.-M., Abergel, C., & Legendre, M. (2020). The DNA methylation landscape of giant viruses. *Nature Communications*, 11(1), 2657. <https://doi.org/10.1038/s41467-020-16414-2>
 38. Woodcock, C. B., Horton, J. R., Zhang, X., Blumenthal, R. M., & Cheng, X. (2020). Beta class amino methyltransferases from bacteria to humans: Evolution and structural consequences. *Nucleic Acids Research*, 48(18), 10034–10044. <https://doi.org/10.1093/nar/gkaa446>
 39. Jeltsch, A. (2001). The cytosine N4-methyltransferase M.PvuII also modifies adenine residues. *Biological Chemistry*, 382(4), 707–710. <https://doi.org/10.1515/bc.2001.084>
 40. Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), 204–220. <https://doi.org/10.1038/nrg2719>
 41. Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, 7(6), 461–465. <https://doi.org/10.1038/nmeth.1459>
 42. Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akesson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, 14(4), 411–413. <https://doi.org/10.1038/nmeth.4189>
 43. Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, 14(4), 407–410. <https://doi.org/10.1038/nmeth.4184>
 44. Dunwell, T. L., & Pfeifer, G. P. (2014). *Drosophila* genomic methylation: New evidence and new questions. *Epigenomics*, 6(5), 459–461. <https://doi.org/10.2217/epi.14.46>
 45. Raddatz, G., Guzzardo, P. M., Olova, N., Fantappié, M. R., Rampp, M., Schaefer, M., Reik, W., Hannon, G. J., & Lyko, F. (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences*, 110(21), 8627–8631. <https://doi.org/10.1073/pnas.1306723110>
 46. Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B., & Nestor, C. E. (2020). No evidence for DNA N (6)-methyladenine in mammals. *Science Advances*, 6(12), eaay3335. <https://doi.org/10.1126/sciadv.aay3335>
 47. Lentini, A., Lagerwall, C., Vikingsson, S., Mjoseng, H. K., Douvlataniotis, K., Vogt, H., Green, H., Meehan, R. R., Benson, M., & Nestor, C. E. (2018). A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nature Methods*, 15(7), 499–504. <https://doi.org/10.1038/s41592-018-0038-7>
 48. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., & Blaxter, M. (2013). Blobology: Exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Frontiers in Genetics*, 4, 237. <https://doi.org/10.3389/fgene.2013.00237>
 49. Delmont, T. O., & Eren, A. M. (2016). Identifying contamination with advanced visualization and analysis practices: Metagenomic approaches for eukaryotic genome assemblies. *PeerJ*, 4, e1839. <https://doi.org/10.7717/peerj.1839>
 50. O'Brown, Z. K., Boulias, K., Wang, J., Wang, S. Y., O'brown, N. M., Hao, Z., Shibuya, H., Fady, P.-E., Shi, Y., He, C., Megason, S. G., Liu, T., & Greer, E. L. (2019). Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. *BMC Genomics*, 20(1), 445. <https://doi.org/10.1186/s12864-019-5754-6>
 51. Krueger, F., Kreck, B., Franke, A., & Andrews, S. R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nature Methods*, 9(2), 145–151. <https://doi.org/10.1038/nmeth.1828>
 52. Yu, M., Ji, L., Neumann, D. A., Chung, D.-H., Groom, J., Westpheling, J., He, C., & Schmitz, R. J. (2015). Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite-sequencing. *Nucleic Acids Research*, 43(21), gkv738. <https://doi.org/10.1093/nar/gkv738>
 53. Schadt, E. E., Banerjee, O., Fang, G., Feng, Z., Wong, W. H., Zhang, X., Kislyuk, A., Clark, T. A., Luong, K., Keren-Paz, A., Chess, A., Kumar, V., Chen-Plotkin, A., Sondheimer, N., Korlach, J., & Kasarskis, A. (2013). Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Research*, 23(1), 129–141. <https://doi.org/10.1101/gr.136739.111>
 54. Clark, T. A., Spittle, K. E., Turner, S. W., & Korlach, J. (2011). Direct detection and sequencing of damaged DNA bases. *Genome Integrity*, 2, 10. <https://doi.org/10.1186/2041-9414-2-10>
 55. Zhu, S., Beaulaurier, J., Deikus, G., Wu, T. P., Strahl, M., Hao, Z., Luo, G., Gregory, J. A., Chess, A., He, C., Xiao, A., Sebra, R., Schadt, E. E., & Fang, G. (2018). Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. *Genome Research*, 28(7), 1067–1078. <https://doi.org/10.1101/gr.231068.117>
 56. Kong, Y., Cao, L., Deikus, G., Fan, Y., Mead, E. A., Lai, W., Zhang, Y., Yong, R., Sebra, R., Wang, H., Zhang, X.-S., & Fang, G. (2022). Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. *Science*, 375(6580), 515–522. <https://doi.org/10.1126/science.abe7489>
 57. Boulias, K., & Greer, E. L. (2022). The adenine methylation debate. *Science*, 375(6580), 494–495. <https://doi.org/10.1126/science.abn6514>
 58. Beh, L. Y., Debelouchina, G. T., Clay, D. M., Thompson, R. E., Lindblad, K. A., Hutton, E. R., Bracht, J. R., Sebra, R. P., Muir, T. W., & Landweber, L. F. (2019). Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. *Cell*, 177(7), 1781–1796.e25. <https://doi.org/10.1016/j.cell.2019.04.028>
 59. Yang, S., Wang, Y., Chen, Y., & Dai, Q. (2020). MASQC: Next generation sequencing assists third generation sequencing for quality control in N6-methyladenine DNA identification. *Frontiers in Genetics*, 11, 269. <https://doi.org/10.3389/fgene.2020.00269>
 60. Takahashi, Y., Shoura, M., Fire, A., & Morishita, S. (2022). Context-dependent DNA polymerization effects can masquerade as DNA modification signals. *BMC Genomics*, 23(1), 249. <https://doi.org/10.1186/s12864-022-08471-2>

61. Guiblet, W. M., Cremona, M. A., Cechova, M., Harris, R. S., Kejnovská, I., Kejnovsky, E., Eckert, K., Chiaromonte, F., & Makova, K. D. (2018). Long-read sequencing technology indicates genome-wide effects of non-B DNA on polymerization speed and error rate. *Genome Research*, 28(12), 1767–1778. <https://doi.org/10.1101/gr.241257>. 118
62. Haute, L. V., Hendrick, A. G., D'Souza, A. R., Powell, C. A., Rebelo-Guimar, P., Harbour, M. E., Ding, S., Fearnley, I. M., Andrews, B., & Minczuk, M. (2019). METTL15 introduces N4-methylcytidine into human mitochondrial 12S rRNA and is required for mitoribosome biogenesis. *Nucleic Acids Research*, 47(19), 10267–10281. <https://doi.org/10.1093/nar/gkz735>
63. Vilkaitis, G., & Klimašauskas, S. (1999). Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. *Analytical Biochemistry*, 271(1), 116–119.
64. Tourancheau, A., Mead, E. A., Zhang, X.-S., & Fang, G. (2021). Discovering multiple types of DNA methylation from bacteria and microbiome using nanopore sequencing. *Nature Methods*, 18(5), 491–498. <https://doi.org/10.1038/s41592-021-01109-3>
65. Li, X., Guo, S., Cui, Y., Zhang, Z., Luo, X., Angelova, M. T., Landweber, L. F., Wang, Y., & Wu, T. P. (2022). NT-seq: A chemical-based sequencing method for genomic methylome profiling. *Genome Biology*, 23(1), 122. <https://doi.org/10.1186/s13059-022-02689-9>
66. Xiong, J., Wang, P., Shao, W.-X., Li, G., Ding, J.-H., Xie, N.-B., Wang, M., Cheng, Q.-Y., Xie, C., Feng, Y.-Q., Ci, W., & Yuan, B.-F. (2022). Genome-wide mapping of N(4)-methylcytosine at single-base resolution by APOBEC3A-mediated deamination sequencing. *Chemical Science*, 13(34), 9960–9972. <https://doi.org/10.1039/d2sc02446b>
67. Yu, L., Zhang, Y., Xue, L., Liu, F., Chen, Q., Luo, J., & Jing, R. (2022). Systematic analysis and accurate identification of DNA N4-methylcytosine sites by deep learning. *Frontiers in Microbiology*, 13, 843425. <https://doi.org/10.3389/fmicb.2022.843425>
68. Chen, W., Yang, H., Feng, P., Ding, H., & Lin, H. (2017). iDNA4mC: Identifying DNA N4-methylcytosine sites based on nucleotide chemical properties. *Bioinformatics*, 33(22), 3518–3523. <https://doi.org/10.1093/bioinformatics/btx479>
69. Ye, P., Luan, Y., Chen, K., Liu, Y., Xiao, C., & Xie, Z. (2017). MethSMRT: An integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. *Nucleic Acids Research*, 45(D1), D85–D89. <https://doi.org/10.1093/nar/gkw950>
70. Manavalan, B., Hasan, M. M., Basith, S., Gosu, V., Shin, T.-H., & Lee, G. (2020). Empirical comparison and analysis of web-based DNA N4-methylcytosine site prediction tools. *Molecular Therapy Nucleic Acids*, 22, 406–420. <https://doi.org/10.1016/j.omtn.2020.09.010>
71. Chen, Y., Wu, Y., Liu, L., Feng, J., Zhang, T., Qin, S., Zhao, X., Wang, C., Li, D., Han, W., Shao, M., Zhao, P., Xue, J., Liu, X., Li, H., Zhao, E., Zhao, W., Guo, X., Jin, Y., ... Zhang, Y. (2019). Study of the whole genome, methylome and transcriptome of *Cordyceps militaris*. *Scientific Reports*, 9(1), 898. <https://doi.org/10.1038/s41598-018-38021-4>
72. Usai, G., Vangelisti, A., Simoni, S., Giordani, T., Natali, L., Cavallini, A., & Mascagni, F. (2021). DNA modification patterns within the transposable elements of the fig (*Ficus carica* L.) genome. *Plants (Basel)*, 10(3), 451. <https://doi.org/10.3390/plants10030451>
73. Liu, Z.-Y., Xing, J.-F., Chen, W., Luan, M.-W., Xie, R., Huang, J., Xie, S.-Q., & Xiao, C.-L. (2019). MDR: An integrative DNA N6-methyladenine and N4-methylcytosine modification database for Rosaceae. *Horticulture Research*, 6, 78. <https://doi.org/10.1038/s41438-019-0160-4>
74. Malone, T., Blumenthal, R. M., & Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *Journal of Molecular Biology*, 253(4), 618–632.
75. Jurkowska, R. Z., & Jeltsch, A. (2016). Mechanisms and biological roles of DNA methyltransferases and DNA methylation: From past achievements to future challenges. In A. Jeltsch & R. Z. Jurkowska (Eds.), *DNA methyltransferases – role and function* (pp. 1–17). Springer International Publishing.
76. Gong, W. (1997). Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Research*, 25(14), 2702–2715. <https://doi.org/10.1093/nar/25.14.2702>
77. Bheemanai, S., Reddy, Y. V. R., & Rao, D. N. (2006). Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochemical Journal*, 399(2), 177–190. <https://doi.org/10.1042/bj20060854>
78. Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., & Shi, Y. (2015). DNA methylation on N6-adenine in *C. elegans*. *Cell*, 161(4), 868–878. <https://doi.org/10.1016/j.cell.2015.04.005>
79. Xiao, C.-L., Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.-Q., Luo, F., Liang, Z., Wang, D.-P., Bo, X.-C., Gu, X.-F., Wang, K., & Yan, G.-R. (2018). N6-methyladenine DNA modification in the human genome. *Molecular Cell*, 71(2), 306–318.e7. <https://doi.org/10.1016/j.molcel.2018.06.015>
80. Woodcock, C. B., Yu, D., Zhang, X., & Cheng, X. (2019). Human HEMK2/KMT9/N6AMT1 is an active protein methyltransferase, but does not act on DNA in vitro, in the presence of Trm112. *Cell Discovery*, 5, 50. <https://doi.org/10.1038/s41421-019-0119-5>
81. Li, W., Shi, Y., Zhang, T., Ye, J., & Ding, J. (2019). Structural insight into human N6amt1-Trm112 complex functioning as a protein methyltransferase. *Cell Discovery*, 5, 51. <https://doi.org/10.1038/s41421-019-0121-y>
82. Kusevic, D., Kudithipudi, S., & Jeltsch, A. (2016). Substrate specificity of the HEMK2 protein glutamine methyltransferase and identification of novel substrates. *Journal of Biological Chemistry*, 291(12), 6124–6133. <https://doi.org/10.1074/jbc.M115.711952>
83. Hamey, J. J., Winter, D. L., Yagoub, D., Overall, C. M., Hart-Smith, G., & Wilkins, M. R. (2016). Novel N-terminal and lysine methyltransferases that target translation elongation factor 1A in yeast and human. *Molecular & Cellular Proteomics*, 15(1), 164–176. <https://doi.org/10.1074/mcp.M115.052449>
84. Drozd, M., Piekarowicz, A., Bujnicki, J. M., & Radlinska, M. (2012). Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Research*, 40(5), 2119–2130. <https://doi.org/10.1093/nar/gkr1039>
85. Beaulaurier, J., Schadt, E. E., & Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. *Nature Reviews Genetics*, 20(3), 157–172. <https://doi.org/10.1038/s41576-018-0081-3>
86. Baum, C., Lin, Y.-C., Fomenkov, A., Anton, B. P., Chen, L., Yan, B., Evans, T. C., Roberts, R. J., Tolonen, A. C., & Ettwiller, L. (2021). Rapid identification of methylase specificity (RIMS-seq) jointly identifies methylated motifs and generates shotgun sequencing of bacterial genomes. *Nucleic Acids Research*, 49(19), e113. <https://doi.org/10.1093/nar/gkab705>
87. Du, J., Johnson, L. M., Jacobsen, S. E., & Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology*, 16(9), 519–532. <https://doi.org/10.1038/nrm4043>
88. Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., & Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, 105(4), 487–497. [https://doi.org/10.1016/s0092-8674\(01\)00324-5](https://doi.org/10.1016/s0092-8674(01)00324-5)
89. Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., & Xu, R.-M. (2014). Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Molecular Cell*, 54(5), 879–886. <https://doi.org/10.1016/j.molcel.2014.04.003>
90. Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature*

- Reviews Molecular Cell Biology*, 20(10), 590–607. <https://doi.org/10.1038/s41580-019-0159-6>
91. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y., & Xu, R.-M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science*, 312(5774), 748–751. <https://doi.org/10.1126/science.1125162>
 92. Lee, J., Thompson, J. R., Botuyan, M. V., & Mer, G. (2008). Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nature Structural & Molecular Biology*, 15(1), 109–111. <https://doi.org/10.1038/nsmb1326>
 93. Jurkowska, R. Z., Qin, S., Kungulovski, G., Tempel, W., Liu, Y., Bashtrykov, P., Stiefelmaier, J., Jurkowski, T. P., Kudithipudi, S., Weirich, S., Tamas, R., Wu, H., Dombrowski, L., Loppnau, P., Reinhardt, R., Min, J., & Jeltsch, A. (2017). H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. *Nature Communications*, 8(1), 2057. <https://doi.org/10.1038/s41467-017-02259-9>
 94. Lyons, D. B., & Zilberman, D. (2017). DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. *Elife*, 6, e30674. <https://doi.org/10.7554/eLife.30674>
 95. Huff, J. T., & Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell*, 156(6), 1286–1297. <https://doi.org/10.1016/j.cell.2014.01.029>
 96. Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463, 237–240. <https://doi.org/10.1038/nature08674>
 97. Tchasovnikarova, I. A., Timms, R. T., Matheson, N. J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M. A., & Lehner, P. J. (2015). Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science*, 348(6242), 1481–1485. <https://doi.org/10.1126/science.aaa7227>
 98. Padeken, J., Methot, S. P., & Gasser, S. M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nature Reviews Molecular Cell Biology*, 23(9), 623–640. <https://doi.org/10.1038/s41580-022-00483-w>
 99. Ponger, L., & Li, W.-H. (2005). Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution*, 22(4), 1119–1128. <https://doi.org/10.1093/molbev/msi098>
 100. Bhattacharyya, M., De, S., & Chakrabarti, S. (2020). Origin and evolution of DNA methyltransferases (DNMT) along the tree of life: A multi-genome survey. *BioRxiv*, 2020.04.09.033167. <https://doi.org/10.1101/2020.04.09.033167>
 101. Ehrlich, M., Gama-Sosa, M. A., Carreira, L. H., Ljungdahl, L. G., Kuo, K. C., & Gehrke, C. W. (1985). DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Research*, 13(4), 1399–1412.
 102. Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C., & Gehrke, C. W. (1986). DNA cytosine methylation and heat-induced deamination. *Bioscience Reports*, 6(4), 387–393.
 103. Hayashi, M., Sugahara, K., Yamamura, A., & Iida, Y. (2021). Evaluation of the properties of the DNA methyltransferase from *Aeropyrum pernix* K1. *Microbiology Spectrum*, 9(2), e0018621. <https://doi.org/10.1128/Spectrum.00186-21>
 104. Amaral Zettler, L. A., Gómez, F., Zettler, E., Keenan, B. G., Amils, R., & Sogin, M. L. (2002). Eukaryotic diversity in Spain's River of Fire. *Nature*, 417(6885), 137–137. <https://doi.org/10.1038/417137a>
 105. Iakovenko, N. S., Smykla, J., Convey, P., Kašparová, E., Kozeretska, I. A., Trokhymets, V., Dykyy, I., Plewka, M., Devetter, M., Duriš, Z., & Janko, K. (2015). Antarctic bdelloid rotifers: Diversity, endemism and evolution. *Hydrobiologia*, 761(1), 5–43. <https://doi.org/10.1007/s10750-015-2463-2>
 106. Shmakova, L., Malavin, S., Iakovenko, N., Vishnivetskaya, T., Shain, D., Plewka, M., & Rivkina, E. (2021). A living bdelloid rotifer from 24,000-year-old Arctic permafrost. *Current Biology*, 31(11), R712–R713. <https://doi.org/10.1016/j.cub.2021.04.077>
 107. Suzuki, A. C., Kagoshima, H., Chilton, G., Grothman, G. T., Johansson, C., & Tsujimoto, M. (2017). Meiofaunal richness in highly acidic hot springs in Unzen-Amakusa National Park, Japan, including the first rediscovery attempt for Mesotardigrada. *Zoological Science*, 34(1), 11–17. <https://doi.org/10.2108/zs160108>
 108. Li, S., Cai, J., Lu, H., Mao, S., Dai, S., Hu, J., Wang, L., Hua, X., Xu, H., Tian, B., Zhao, Y., & Hua, Y. (2019). N⁴-cytosine DNA methylation is involved in the maintenance of genomic stability in *Deinococcus radiodurans*. *Frontiers in Microbiology*, 10, 1905. <https://doi.org/10.3389/fmicb.2019.01905>
 109. Douki, T., Meador, J. A., Bérard, I., & Wack, A. (2015). N4-methylation of cytosine drastically favors the formation of (6-4) photoproducts in a TCG context. *Photochemistry and Photobiology*, 91(1), 102–108. <https://doi.org/10.1111/php.12365>
 110. Yamamoto, J., Oyama, T., Kunishi, T., Masutani, C., Hanaoka, F., & Iwai, S. (2013). A cyclobutane thymine–N4-methylcytosine dimer is resistant to hydrolysis but strongly blocks DNA synthesis. *Nucleic Acids Research*, 42(3), 2075–2084. <https://doi.org/10.1093/nar/gkt1039>
 111. Engel, J. D., & Von Hippel, P. H. (1978). Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. *Journal of Biological Chemistry*, 253(3), 927–934.
 112. Iurgaitis, A. P., Butkus, V. V., Klimasauskas, S. I., & Janulaitis, A. A. (1988). Effect of N4-methylcytosine and 5-methylcytosine on the stability of the DNA helix. *Bioorganicheskaya Khimiya*, 14(2), 158–165.
 113. Butkus, V., Klimašauskas, S., Petrauskienė, L., Maneliene, Z., Janulaitis, A., Minchenkova, L. E., & Schyolkina, A. K. (1987). Synthesis and physical characterization of DNA fragments containing N4-methylcytosine and 5-methylcytosine. *Nucleic Acids Research*, 15(20), 8467–8478. <https://doi.org/10.1093/nar/15.20.8467>
 114. Ngo, T. T. M., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A., & Ha, T. (2016). Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nature Communications*, 7(1), 10813. <https://doi.org/10.1038/ncomms10813>
 115. Pérez, A., Castellazzi, C. L., Battistini, F., Collinet, K., Flores, O., Deniz, O., Ruiz, M. L., Torrents, D., Eritja, R., Soler-López, M., & Orozco, M. (2012). Impact of methylation on the physical properties of DNA. *Biophysical Journal*, 102(9), 2140–2148. <https://doi.org/10.1016/j.bpj.2012.03.056>
 116. Gladyshev, E. A., & Arkhipova, I. R. (2010). Genome structure of bdelloid rotifers: Shaped by asexuality or desiccation? *Journal of Heredity*, 101(Suppl 1), S85–S93. <https://doi.org/10.1093/jhered/esq008>
 117. Gladyshev, E. A., & Arkhipova, I. R. (2010). A subtelomeric non-LTR retrotransposon Hebe in the bdelloid rotifer *Adineta vaga* is subject to inactivation by deletions but not 5' truncations. *Mobile DNA*, 1(1), 12. <https://doi.org/10.1186/1759-8753-1-12>
 118. Rodriguez, F., Kenefick, A., & Arkhipova, I. (2017). LTR-retrotransposons from bdelloid rotifers capture additional ORFs shared between highly diverse retroelement types. *Viruses*, 9(4), 78.
 119. Arkhipova, I. R., Yushenova, I. A., & Rodriguez, F. (2013). Endonuclease-containing Penelope retrotransposons in the bdelloid rotifer *Adineta vaga* exhibit unusual structural features and play a role in expansion of host gene families. *Mobile DNA*, 4(1), 19. <https://doi.org/10.1186/1759-8753-4-19>
 120. Alexeeva, M., Guragain, P., Tesfahun, A. N., Tomkuvienė, M., Arshad, A., Gerasimaitė, R., Rukšėnaitė, A., Urbanavičiūtė, G., Björås, M., Laerdahl, J. K., Klungland, A., Klimašauskas, S., & Bjelland, S. (2018). Excision of the doubly methylated base N⁴,5-dimethylcytosine from DNA by *Escherichia coli* Nei and Fpg proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1748), 20170337. <https://doi.org/10.1098/rstb.2017.0337>
 121. Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, 13(8), 335–340. [https://doi.org/10.1016/s0168-9525\(97\)01181-5](https://doi.org/10.1016/s0168-9525(97)01181-5)

122. Rothi, M. H., & Greer, E. L. (2022). From correlation to causation: The new frontier of transgenerational epigenetic inheritance. *BioEssays*, 45(1), e2200118. <https://doi.org/10.1002/bies.202200118>
123. Sadler, K. C. (2022). Epigenetics across the evolutionary tree: New paradigms from non-model animals. *BioEssays*, 45(1), e2200036. <https://doi.org/10.1002/bies.202200036>
124. Aravind, L., Abhiman, S., & Iyer, L. M. (2011). Natural history of the eukaryotic chromatin protein methylation system. *Progress in Molecular Biology and Translational Science*, 101, 105–176. <https://doi.org/10.1016/B978-0-12-387685-0.00004-4>
125. Wang, S., Xie, H., Mao, F., Wang, H., Wang, S., Chen, Z., Zhang, Y., Xu, Z., Xing, J., Cui, Z., Gao, X., Jin, H., Hua, J., Xiong, B., & Wu, Y. (2022). N(4)-acetyldeoxycytosine DNA modification marks euchromatin regions in *Arabidopsis thaliana*. *Genome Biology*, 23(1), 5. <https://doi.org/10.1186/s13059-021-02578-7>
126. Zhou, J., Wang, X., Wei, Z., Meng, J., & Huang, D. (2022). 4acCPred: Weakly supervised prediction of N (4)-acetyldeoxycytosine DNA modification from sequences. *Molecular Therapy Nucleic Acids*, 30, 337–345. <https://doi.org/10.1016/j.omtn.2022.10.004>
127. Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Müller, U., Spruijt, C. G., ... Carell, T. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature Chemical Biology*, 10(7), 574–581. <https://doi.org/10.1038/nchembio.1532>
128. Nguyen Ba, A. N., Pogoutse, A., Provart, N., & Moses, A. M. (2009). NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, 10, 202. <https://doi.org/10.1186/1471-2105-10-202>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Arkhipova, I. R., Yushenova, I. A., & Rodriguez, F. (2023). Shaping eukaryotic epigenetic systems by horizontal gene transfer. *BioEssays*, 45, e2200232. <https://doi.org/10.1002/bies.202200232>