ANNUAL REVIEWS

Annual Review of Microbiology

Beyond Restriction Modification: Epigenomic Roles of DNA Methylation in Prokaryotes

Brian P. Anton and Richard J. Roberts

New England Biolabs, Ipswich, Massachusetts 01938, USA; email: anton@neb.com, roberts@neb.com

Annu. Rev. Microbiol. 2021. 75:129–49

First published as a Review in Advance on July 27, 2021

The *Annual Review of Microbiology* is online at micro.annualreviews.org

https://doi.org/10.1146/annurev-micro-040521-035040

Keywords

methylation, restriction modification, methyltransferase, phase variation, epigenetic, gene regulation

Abstract

The amount of bacterial and archaeal genome sequence and methylome data has greatly increased over the last decade, enabling new insights into the functional roles of DNA methylation in these organisms. Methyltransferases (MTases), the enzymes responsible for DNA methylation, are exchanged between prokaryotes through horizontal gene transfer and can function either as part of restriction-modification systems or in apparent isolation as single (orphan) genes. The patterns of DNA methylation they confer on the host chromosome can have significant effects on gene expression, DNA replication, and other cellular processes. Some processes require very stable patterns of methylation, resulting in conservation of persistent MTases in a particular lineage. Other processes require patterns that are more dynamic yet more predictable than what is afforded by horizontal gene transfer and gene loss, resulting in phase-variable or recombination-driven MTase alleles. In this review, we discuss what is currently known about the functions of DNA methylation in prokaryotes in light of these evolutionary patterns.

Contents

INTRODUCTION 130
PATTERNS OF EVOLUTION IN PROKARYOTIC METHYLATION 132
PHASE VARIATION 134
TRD CASSETTE EXCHANGE 136
CONSERVATION AND FUNCTION OF TARGET SITE–DIVERSE MTase LOCI 136
ACQUISITION OF NOVEL TARGET SITES BY HORIZONTAL GENE TRANSFER AND MUTATION 137
PERSISTENT MTases 139
- Dam 140
- CcrM 140
- Dcm 141
- CamA 141
- *Burkholderia* MTases 141
- CtsM 142
- Other Examples 142

CONCLUSIONS AND FUTURE DIRECTIONS 142

INTRODUCTION

Bacteria have played key roles in our understanding of DNA methylation in all domains of life. It was in the DNA of a bacterium, *Mycobacterium tuberculosis*, that the first methylated nucleobase was discovered in 1925 (63). It was likewise in the extract of a bacterium, *Escherichia coli* W, that enzymatic methylation of a DNA polymer from *S*-adenosylmethionine (SAM) was first demonstrated in 1963 (45). As we now know, DNA methylation is an epigenetic signal capable of altering the interpretation of genetic information without change to the DNA sequence itself. In eukaryotes, DNA methylation is widespread. Notably, in mammals it is conferred by a broadly conserved apparatus and has a role in cancer (12, 35, 42), all of which has led to much progress in elucidating its functional roles. In bacteria and archaea, which are more genetically diverse, we observe numerous DNA methylation systems independently acquired and developed by different taxonomic groups. Although many of these systems remain understudied from the perspective of biological function, the rapid accumulation of prokaryotic genome and methylome data over the last decade has significantly advanced our understanding.

Most enzymatic DNA methylation occurs after DNA replication and is catalyzed by DNA methyltransferases (MTases) that transfer a methyl group from SAM to specific base positions within the DNA polymer. All known DNA MTases adopt a common protein fold (76, 104), and conservation of structural and catalytic elements in these enzymes has manifested as a series of characteristic protein sequence motifs, making the de novo identification of DNA MTase–encoding genes in genome sequences relatively straightforward (74, 96). DNA methylation has been observed in at least 93% of prokaryotic genomes examined (17), making these enzymes nearly ubiquitous. As for biological function, the consequences of DNA methylation to an organism derive from three properties: the enzymatic product of the MTase, the spatial pattern of methylation, and the temporal pattern of methylation.

There are three known enzymatic DNA methylation products: 5-methylcytosine (m5C) (discovered in 1925), N^6-methyladenine (m6A) [identified in 1955 (31)], and N^4-methylcytosine (m4C) [identified in 1983 (59)]. All three of these products occur in prokaryotes, while at present only the first two have been identified in eukaryotes (3, 62). These modifications do not affect Watson-Crick base-pairing but rather protrude into the major groove, where they affect the interaction of proteins such as transcription factors and replication and repair enzymes with the DNA double helix (65). The methyl group may affect protein interactions negatively, typically through steric hindrance, or positively, by recruiting specialized proteins evolved to recognize and bind methylated bases. Although most functional consequences of DNA methylation in prokaryotes are believed to proceed from this property, the chemistry of the methylated base itself is also important. For example, spontaneous deamination of m5C results in the natural base thymine, which is less likely to be repaired than unnatural bases, leading to excess C→T transitional mutations (27). In eukaryotes, m5C can also act as a substrate for oxidation by TET enzymes, leading to moieties that are more chemically active (52, 56, 69, 116). As TET-like genes have been identified in bacteriophages and bacteria (57, 58), it appears likely that such systems will eventually be characterized in prokaryotes as well.

Most DNA MTases methylate a given base only when it occurs in a specific DNA sequence context, known as the target site (TS), recognition site, or target motif. Critically, the TS specifies the spatial pattern of methylation throughout the DNA molecule. TSs are short (typically two to ten specified bases), may be palindromic or nonpalindromic, and may contain unspecified or degenerate positions. The TS sequence is determined by one or more target recognition domains (TRDs), which are usually part of the MTase itself but sometimes occur on a separate, specificity-determining protein with which the MTase directly interacts. TRDs typically exhibit far less sequence conservation than the catalytic and SAM-binding regions of the MTase.

DNA methylation marks are chemically stable. In eukaryotes, DNA methylation can be removed through TET oxidation and repair pathways (52), but such active demethylation has not been observed in prokaryotes. Instead, demethylation occurs passively, through DNA replication. Newly synthesized strands are unmethylated, so the temporal pattern of methylation is largely governed by the time gap between DNA synthesis and remethylation. This in turn is determined by the expression level of the MTase in the cell and its access to the DNA.

In prokaryotes, DNA MTases often form part of restriction-modification (RM) systems, which are commonly thought of as primitive immune systems. They comprise a restriction endonuclease (REase) activity, which degrades foreign DNA by introducing double-strand breaks, and an MTase activity, which protects the host's own chromosome from degradation through steric hindrance of the REase at its binding sites. RM-associated DNA methylation patterns therefore distinguish self from nonself DNA. RM systems, and DNA MTases in general, are widespread throughout the bacterial and archaeal domains due to not the vertical inheritance of a common ancestral MTase but rather horizontal gene transfer (HGT).

RM systems are classified into four major types based on protein stoichiometry and overall mode of action (100). Type I systems require three proteins: an REase and motor protein (R; encoded by *hsdR*), an MTase (M; encoded by *hsdM*), and a specificity subunit (S; encoded by *hsdS*) that contains the TRDs and therefore determines the TS. Methylation, performed by an M_2S protein complex, occurs on both strands of a bipartite TS. Cleavage, performed by an R_2M_2S complex, occurs at random distances from the TS, as the ATP-dependent motor function of R loops DNA through the complex, with cleavage occurring when translocation is interrupted. Type III systems require two proteins: an REase (Res) and an MTase (Mod), with specificity governed solely by the Mod protein. Type III protein products also act as complexes, Mod_2 for methylation and Res_1Mod_2 for cleavage (21, 49). Methylation occurs on one strand of a nonpalindromic site, and

Table 1 Number of prokaryotic methyltransferase genes in REBASE[a]

RM system type	Part of complete RM system	Orphan methyltransferase
I	22,762	648
IIG	12,160	Not applicable
II (all other subtypes)	14,050	29,138
III	8,801	578

Abbreviation: RM, restriction modification.

[a]Based on data from 23,535 complete genomes and includes some unpublished data. Accessed October 11, 2020, at **http://rebase.neb.com/rebase/statmlist.html**.

cleavage occurs at a fixed distance 20–30 bp away. Type IV systems recognize and cleave methylated TSs and comprise an REase and occasionally an MTase that is necessarily nonfunctional.

Type II systems are a disparate group that includes any RM system in which cleavage occurs at a fixed position within or close to its TS (100). Type IIG, Type IIB, and Type IIC systems, some of which have likely evolved from Type I systems (18, 80), encode a single polypeptide that performs both REase and MTase functions. Most Type II systems, however, comprise REase and MTase functions that are performed by independent proteins that do not physically interact. By necessity, the two proteins must recognize the same TS, or at least the MTase must methylate all instances of the REase TS to prevent frequent damage to the host chromosome. This convergence may occur through a combination of coevolution, reduction from more complex systems, and gene reassortment to bring compatible REase and MTase genes together to form a functional RM system (18, 53).

This process of gene reassortment relies on a metagenomic pool of Type II RM systems and component genes with different specificities from which to build new systems. While unpaired Type II REases are detrimental to the cell, unpaired MTases are less so. Consequently, in sequenced genomes there appears to be a large number of so-called orphan MTases, which have no apparent cognate REase gene (**Table 1**). It must be noted, however, that genes encoding Type II REases are often difficult to identify by sequence similarity (93, 125), and so it is possible that some apparently orphan MTases may have unrecognized REase partners. Over the last several decades, it has become apparent that at least some truly orphan MTases have been co-opted to perform novel functions.

Bacteria and archaea have evolved many ways to take advantage of DNA methylation, and whether it occurs with or without a cognate REase may be an artificial distinction. Many methylation-related functions involve the alteration of gene expression by affecting the binding of transcription factors at upstream gene-regulatory regions, but methylation has other uses as well, such as the marking of parental DNA strands and environmental DNA uptake. In the remainder of this review, we provide a brief overview of present knowledge of prokaryotic DNA methylation, and we direct readers to other excellent reviews for further detail.

PATTERNS OF EVOLUTION IN PROKARYOTIC METHYLATION

We can infer a great deal about the function of a particular pattern of DNA methylation by observing, through either cell-to-cell variation or comparative genomics, the timescale on which it changes. This change must take one of two forms: gain or loss of methylation at a particular TS (governed by what we refer to as an *A*/OFF switch to indicate the gain or loss of methylation at TS *A*) or alteration of the TS (determined by what we refer to as an *A*/*B* switch to indicate a change from methylation at TS *A* to methylation at TS *B*). Different types and timescales of change are

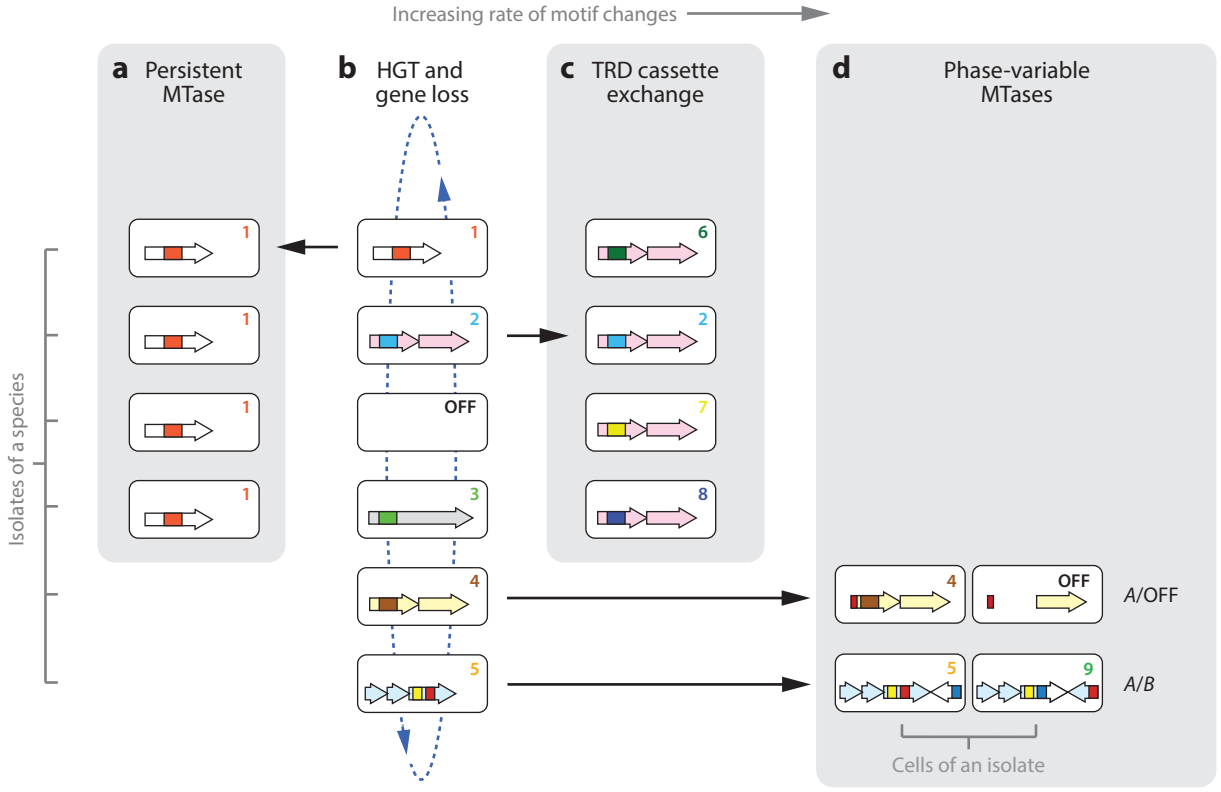

Figure 1

Schematic overview of the evolution of prokaryotic DNA methylation patterns, with four evolutionary mechanisms. Wide arrows represent genes, with orthologous genes at the same locus shown in the same color. Colored boxes within genes represent TRDs, with different colors representing different TSs or TS half-sites and colored numerals representing the TS associated with the MTase. Only one MTase or RM system is shown per cell for simplicity. Each row in each panel represents an isolate within a species, although other taxonomic scales are possible. (*a*) A lineage-specific, or persistent, orphan MTase, with an identical motif in each isolate. (*b*) The diversity of MTases and RM systems created by HGT between isolates and the exogenous metagenomic pool. The word OFF indicates that no active DNA MTase is present. The blue dashed oval with arrowheads conceptually represents the HGT process. Black arrows show that acquisition by HGT is often the starting point for the development of the evolutionary paradigms in the other panels. (*c*) A cassette-variable Type III RM system, where the locus remains fixed but the TRD is exchanged by HGT and recombination. (*d*) Two examples of phase-variable MTases, where the strain itself exhibits diversity: (*top*) a Type III RM system with a phase-variable simple sequence repeat (*red*) acting as an *A*/OFF switch and (*bottom*) a Type I RM system with an invertible exchange of the second TRD of the specificity subunit gene acting as an *A*/*B* switch (in this case toggling between TSs 5 and 9). (Because two TRDs contribute to the TSs of these Type I RM systems, the colors of numerals 5 and 9 are blends of the colors of active TRDs.) Abbreviations: HGT, horizontal gene transfer; MTase, methyltransferase; RM, restriction modification; TRD, target recognition domain; TS, target site.

achieved by different evolutionary mechanisms (**Figure 1**), which can be summarized as follows, roughly in order from most rapidly changing to most stable: (*a*) phase-variable expression, (*b*) recombination of TRDs, (*c*) gain and loss of MTase genes through HGT, (*d*) mutation of TRDs, and (*e*) vertical inheritance of MTases with unchanged specificity.

When methylation of a specific TS becomes advantageous, the MTase (including its TRD) will be retained through purifying selection, resulting in persistence. This type of methylation typically affects the internal workings of the cell, such as mismatch repair, DNA replication, and the predictable regulation of specific genes. When methylation of a diverse suite of alternative TSs becomes advantageous, any number of more rapid mechanisms of change can be employed,

including cassette exchange and phase variation. This type of methylation is most commonly involved with environmental interaction and adaptation.

The classic cellular defense function of RM systems may occupy the intermediate temporal space, ensuring the continued HGT of DNA MTases, whether as orphans or as part of RM systems, and the mutation of TRDs to change from one TS to another. Protection from bacteriophage infection is a benefit conferred by RM systems, but a transient one. At each intrusion event, DNA methylation and restriction compete to act first on the foreign DNA. In most cases, restriction acts first, the target is destroyed, and infection is prevented. However, in the rare cases where DNA methylation acts first, not only does the RM system fail to protect the cell from infection but the subsequently created phage particles are methylated and therefore immune to restriction by other cells in the population carrying the same RM system(s). There is thus a constant interplay of evolutionary forces acting on prokaryotic DNA MTases through HGT, purifying, and diversifying selection.

PHASE VARIATION

Phase variation is a strategy to generate genotypic and phenotypic diversity in a bacterial population in the absence of selection. This is achieved through any of several stochastic and reversible genetic mechanisms, such as slipped-strand mispairing at simple sequence repeats (SSRs) and high-frequency domain shuffling via repeat sequences. Phase-variable switches, whether *A*/OFF or *A*/*B*, are hypermutable, with SSR changes occurring at a rate several orders of magnitude faster than that of base mutation (83). The result is a cell population with multiple states at the affected loci, enabling it to quickly respond and adapt to changing environmental conditions. Pathogenic bacteria, for example, often exploit phase variability to evade host immune responses (10).

While many of the genes under phase-variable control—encoding adhesins, pili, iron receptors, and lipopolysaccharides (64, 118, 123, 133)—relate directly to environmental interaction, a significant number of genes for DNA MTases are also phase variable, as described in several recent reviews (10, 29, 106). The genome-wide change in methylation pattern that results from the phase-variable expression of a single MTase can alter the expression of many genes simultaneously, and the suite of genes under this kind of control has been termed a phasevarion (phase-variable regulon) (115). Most phase-variable MTases form part of a Type I, Type III, or Type IIC RM system, in which the REase cannot function in the absence of the MTase and the specificity of both is governed by a shared TRD. Three types of phase-variable mechanisms that control DNA methylation have been observed: SSR expansion and contraction as either an *A*/OFF or *A*/*B* switch, and recombination at repeat sequences as an *A*/*B* switch.

SSR expansion and contraction is a strategy employed by all types of RM systems. In Type III *mod* genes, SSR tracts are most commonly found near the 5′ end and function exclusively as *A*/OFF switches. Gain or loss of repeat units alters the reading frame of that portion of the gene downstream of the SSR, resulting in either a functional MTase (methylating TS *A*) or a truncated, nonfunctional polypeptide (the OFF state). At least 17.4% of Type III *mod* genes have SSRs characteristic of phase-variable expression, and differences in both the sequence and location of the repeat tracts suggest that this phenomenon has arisen in this group at least 25 times independently (11). Potentially phase-variable SSRs are also present in at least 2% of Type I *hsdM* genes, where they occur exclusively at the 5′ ends and are believed to function as *A*/OFF switches in an analogous fashion (8).

SSRs have also been observed in at least 7.9% of Type I *hsdS* genes, where the examples studied function exclusively as *A*/*B* switches (8). Type I RM systems methylate bipartite TSs comprising two half-sites (each typically two to five base pairs) separated by a nonspecific spacer region

(typically four to eight base pairs). A typical S protein contains two globular TRDs (TRD1 and TRD2), each responsible for recognizing one half-site, joined by two long, antiparallel alpha helices whose structure is related to the length of the spacer region (22). All known SSRs within *hsdS* genes are located in the conserved central region of the gene, which encodes the alpha-helical spacer (8). A full-length S protein recognizes a nonpalindromic TS with the half-sites determined by TRD1 and TRD2, while a truncated S protein resulting from an SSR-mediated frameshift recognizes a palindromic TS with both half-sites determined by TRD1 (8). A related arrangement has been observed in atypical *hsdS* genes that encode three TRDs (TRD1, TRD2a, and TRD2b) instead of the usual two, with the SSR occurring between TRD2a and TRD2b. When TRD2b is out of frame with the others, an S protein containing TRD1 and TRD2a is expressed. When TRD2b is in frame with the others, a nonstandard S protein containing all three TRDs is expressed, but only the N-terminal and C-terminal TRDs (TRD1 and TRD2b) determine the TS (8). Exactly how this three-TRD protein folds to exclude TRD2a from participating in the TS is unknown.

A small number of examples of SSR switches have been described in Type IIC RM systems, in which the MTase and REase functions are determined by the same polypeptide. In *Campylobacter jejuni* 11168, a polyguanidine tract within the MTase domain serves as an *A*/OFF switch in *cj0031*, which encodes the protein CjeNV (4). Two orthologs from *Helicobacter pylori* have been described—encoding Hpy99XIV in strain J99 and HpyAXVI in strain 26695—both of which contain two polycytidine tracts (68). In both orthologs, the upstream tract functions as an *A*/OFF switch in similar fashion to *cj0031*, while the downstream tract functions as an *A*/*B* switch (68). Together, the two tracts function as three-way *A*/*B*/OFF switches. The BcgI-like Type IIC system Hpy99XXII (TCAN$_6$TRG) comprises an RM protein and a separate S protein. A polyguanidine tract within the S gene appears to function as an *A*/OFF switch in this system (68), in contrast to Type I *hsdS* examples.

Phase variation of standard Type II RM systems, in which the MTase and REase proteins act separately, has to our knowledge been shown in only a single case, that of M.HpyAIV, which methylates GANTC. This MTase was found in 60% of a group of Swedish clinical *H. pylori* isolates. In that study, activity of the MTase was strongly correlated with the length of a polyadenosine tract in the gene, but the activity of the cognate REase was not investigated (110). It remains an open question as to how the REase activity is regulated to accommodate the OFF switching of the MTase, but there is some evidence of coregulation at the transcriptional level (110). Although the length of the SSR tract varies from isolate to isolate, the timescale of change may be slower than observed for Type I and Type III RM systems, since no variation was observed within a single-colony isolate (110). As only a small number of examples have been described, Type IIC and Type II RM systems would likely benefit from the type of large-scale search for potentially phase-variable SSRs that was undertaken for Type I and Type III systems (7, 11).

In addition to SSRs, Type I RM systems also use another type of *A*/*B* switch that involves rapid recombination at repeat sequences to exchange TRDs between full and partial *hsdS* genes. When TRDs are shuffled into and out of the actively transcribed *hsdS*, the TS is altered at one or both half-sites. A single described case, the *tvr* (SpnIV) locus of *Streptococcus pneumoniae*, uses an excision-integration system based on direct repeats that involves a circular intermediate stabilized by a toxin-antitoxin system (70). All other such systems presently known use inverted repeats (IRs) to exchange TRDs by rapid inversion, a somewhat simpler approach. One well-characterized such system, the *ivr* locus (SpnIII) from *S. pneumoniae*, uses three sets of IRs and five different TRDs to generate six possible *hsdS* alleles, designated A–F, each with a different TS (75). Systems of this type have also been described in *Mycoplasma pulmonis* (8, 11), *Bacteroides fragilis* (7, 14, 54), *Mannheimia haemolytica* (7), *Streptococcus suis* (124), *Lactobacillus salivarius* (28), *Listeria monocytogenes* (7, 29, 33), and others, and overall, about 3.9% of all Type I RM systems have sequence features

characteristic of IR-mediated recombination (7). The recombinase required for inversion is often located within the *hsd* locus and within the inverted segment itself, as in the cases of *S. pneumoniae* and *M. haemolytica*. In other cases, such as the HsvR recombinase of *M. pulmonis*, the recombinase gene is distal to the *hsd* locus where it acts (109). These systems have been classified into five groups based on the type of recombinase that is used (54).

Nearly all of the phase-variable MTases described to date, whether Type I, II, or III, have a cognate REase, and so an intriguing question is, What is the role of the REase? One might expect that orphan Type II MTases, which are prevalent in prokaryotic genomes, might just as easily evolve phase-variable SSR mechanisms, but no such cases have been observed, suggesting a specific biological role either for the REase or for the overall RM system architecture. While phase variability has independently evolved in a minority, albeit a significant minority, of RM systems, once it does evolve, it appears to stabilize the retention of the system. Without a phase-variable switch, the logical OFF switch for a DNA MTase is loss of the gene.

TRD CASSETTE EXCHANGE

Besides those described above that are deemed phase variable, other types of recombination events have been described, particularly in Type III *mod*, Type I *hsdS*, and Type IIB S-encoding genes, all of which have conserved elements involved with catalysis, structure, or protein-protein interaction. When conserved at the DNA level, these regions promote recombination at TRD regions between them, essentially leaving the gene framework intact but exchanging the TRD(s) in a cassette-like fashion. In Type I *hsdS* genes, TRDs in the same relative position can be easily exchanged, presumably by homologous recombination (40). However, TRDs in different positions can also be exchanged via short direct-repeat flanking sequences, a process referred to as DoMo (domain movement) that may work by either homologous or site-specific recombination (40). Type III *mod* genes contain a single TRD that in some cases can be switched between different alleles, many of which are also phase variable through SSR *A*/OFF mechanisms (106). For example, the *modH* locus of *H. pylori* has 17 known alleles (114), and similar examples can be found in *Haemophilus*, *Neisseria*, and *Moraxella* species (16, 43, 102, 105, 113, 117) (**Table 2**). Each allele recognizes a different TS and controls a distinct phasevarion (43). The evolution of these cassette systems is complex, involving at times recombination between species, between isolates, between alleles, or within alleles (16, 41, 43).

CONSERVATION AND FUNCTION OF TARGET SITE–DIVERSE MTase LOCI

TRD cassette recombination and phase-variable MTases generate diverse methylation patterns among isolates of the same species and individual cells of the same isolate or strain, respectively. Counterintuitively, this variability is itself conserved. While many MTase and other component genes of RM systems are not conserved between isolates of the same species, those MTase genes that have evolved phase variability or cassette exchange often persist within species or sublineages thereof (**Table 2**). This implies selection for a particular function or functions in which diverse and readily alterable methylation patterns are advantageous, such as cellular defense (i.e., protection from diverse bacteriophages) and environmental response (i.e., generation of phenotypic diversity). The former of these functions requires the presence of a cognate REase, while the latter does not, or at least not obviously. The cellular defense function is beyond the scope of this review, so we focus on the environmental and epigenetic roles here.

Any given DNA MTase is likely to modify hundreds or thousands of sites throughout the genome, most of which will have little or no effect on gene regulation. A small number, however,

Table 2 Examples of cassette-variable Type III methyltransferase alleles

System	Alleles (*n*)	Species	Example	Frequency (percentage of genomes)[a]	Reference(s)
modH	17	*Helicobacter pylori*	M.Hpy99XXI	100	114
modA	21	*Haemophilus influenzae*	M.HindVI	100	9, 43, 117
		Neisseria meningitidis	M.Nme214I	100	
		Neisseria gonorrhoeae	M.NgoWNIV	100	
modB	6	*Neisseria meningitidis*	M.Nme18I	84	117
		Neisseria gonorrhoeae	M.Ngo3502I	100	
modD	7	*Neisseria meningitidis*	M.Nme15123I	80	117
		Neisseria gonorrhoeae	M.NgoWFI	95	
modM	6	*Moraxella catarrhalis*	M.Mca195I	69	16
modN	2	*Moraxella catarrhalis*	M.Mca23246IIP	0[b]	16
modO	6	*Moraxella catarrhalis*	M.McaC031ORF1728P	0[b]	16

[a]Percentage of completely sequenced genomes of this species encoding the gene in question, regardless of which target recognition domain is present. Number of genomes: *H. pylori*, 249; *H. influenzae*, 72; *N. meningitidis*, 115; *N. gonorrhoeae*, 63; *M. catarrhalis*, 16. Data from REBASE (**http://rebase.neb.com**).
[b]These genes occur only in *M. catarrhalis* RB2/3 strains (16), for which there are no complete genomes available in REBASE.

will affect the expression of a subset of genes (the regulon or phasevarion, depending on the timescale). The example of *modA11* in *Neisseria meningitidis* shows these effects can be either direct or indirect: Of 285 genes differentially regulated between the *modA11* ON and OFF states (113), only 26 were directly regulated through methylation of the upstream intergenic region, while the remainder were likely indirectly regulated through perturbation of regulatory factor expression or key biological processes (61).

Furthermore, it may be that epigenetic regulation of only a subset of the regulon or phasevarion confers any selective or phenotypic advantage, and determining which regulated genes are functionally important and which are coincidental is challenging. Comparative genomic analysis of methylation sites across sets of isolates with a common MTase allele may shed light on this problem, but to our knowledge such studies have not yet been carried out. In the meantime, studies have focused on the sets of functions attributed to phasevarions and regulons, noting particularly those pertaining to environmental interaction, immunological response, or virulence (9, 115).

ACQUISITION OF NOVEL TARGET SITES BY HORIZONTAL GENE TRANSFER AND MUTATION

The establishment and expansion of these kinds of multi-allele systems require not only the shuffling of TRDs already in the system but also the creation or recruitment of new TRDs to introduce novel alleles. TRDs recognizing a given TS can be acquired by HGT or evolve from TRDs recognizing other TSs. HGT in particular has long been acknowledged as critical to the dispersion of RM systems and their component genes based on such evidence as codon usage discrepancy and incongruence between gene and species trees (51, 60). RM systems and orphan MTases are both associated with mobile genetic elements such as plasmids, prophages, integrons, and transposable elements, or with fragments thereof (5, 15, 39, 55, 66, 101, 119). However, there are vehicular preferences, with virulent bacteriophages and plasmids more likely to include orphans than full RM systems (91). One plasmid-borne orphan MTase, M.EcoGIX, may even be mechanistically involved with HGT and/or plasmid replication itself, as it is highly

strand-specific (34, 128). In addition, DNA MTases, or at least those belonging to RM systems, are overrepresented in naturally competent organisms (91).

They appear to be lost as frequently as they are gained, creating an ever-changing repertoire of DNA MTases in prokaryotic genomes. Genome comparison studies within such taxonomic groups as *H. pylori*, *N. meningitidis*, and *Microcystis* show significant isolate-to-isolate differences in RM system content (20, 120) and DNA methylation patterns (68, 130). Despite the relatively rapid gene loss, DNA MTase gene sequences appear to be under purifying selection, suggesting alternating periods of positive and relaxed selection (91). Various functions have been proposed for these exchanged RM systems and orphan MTases, which have been co-opted neither for allelic/phase-variable systems nor for the more stable, moonlighting functions described in the next section. Most of these functions (reviewed in 122) involve the discrimination of self from nonself DNA.

The most traditional of such discrimination functions is cellular defense. Defense mechanisms based on DNA methylation include not only RM systems but also the more recently discovered BREX systems (46), which prevent bacteriophage infection but do not employ a companion REase. Their mode of action is not well understood, but it is believed that the MTase (PglX) and a companion alkaline phosphatase (PglZ) act as a toxin-antitoxin pair, and it has been demonstrated that DNA methylation by PglX confers immunity (48). It is possible that other orphan MTases have similar roles in marking self DNA but partner with other unknown system components to defend against attacking bacteriophages (see, for example, 30).

Methylation-based defense mechanisms limit not only bacteriophage infection but also the HGT of other mobile elements, functioning more generally as immigration control agents (85, 108). Methylation patterns limit the exchange of genetic information between cells with incompatible patterns, creating the genetic isolation needed for the establishment of new biotypes and species (85, 122). Conversely, Oliveira and coworkers (92) have shown that organisms with compatible suites of Type II RM systems undergo more frequent genetic coexchange via HGT and homologous recombination regardless of the evolutionary distance between them. The frequent turnover of DNA MTases through HGT causes the definitions of self and nonself to continuously change, leading to very complex, time-dependent networks of genetic flux. The rate, nature, and ramifications of this change likely differ widely between species.

Another explanation for the frequent HGT of RM systems is that they act as selfish elements or addictive modules, promoting their own survival and propagation (67, 86). This concept applies primarily to Type II systems, where the REase and MTase functions act independently. Whenever the genetic system (or the MTase gene alone) is lost, either by loss of an encoding plasmid or by genetic inactivation, transcription ceases and the remaining REase and MTase molecules in the cell become progressively diluted. Eventually, there is not enough MTase to prevent double-strand breaks in the chromosome by the remaining REase and the cell dies. Consequences of this include plasmid retention and potentially genomic island stabilization (122). Despite this selfish behavior, loss of Type II RM systems clearly does occur, and one way it can occur is loss or inactivation of the REase gene alone (71, 107), which may create a pool of orphan Type II MTase genes.

The methylome of a cell can change not only through the gain and loss of MTases acting at specific TSs but also through the evolution of MTase sequences. While the TSs of many MTases are determined by true, discrete TRDs (enabling the cassette-like recombination of Type I and Type III RM systems described above), those of some others are determined by discontinuous sequence elements such as positively charged protein loops (74, 97, 98). In the case of Type II MTases, protein sequences with greater than about 50% sequence identity tend to recognize the same TS (92). However, in the case of Type I, Type III, and certain Type IIG enzymes, alteration of one or a very small number of amino acids within the TRD is sufficient to change the TS (1, 92, 111). In the case of the Type IIG RM protein MmeI, this has even led to the rational engineering

of specificity (23, 81). Thus, the degree of genetic alteration required to change the methylated TS, and therefore the frequency of such TS changes, is likely to vary widely from case to case.

PERSISTENT MTases

Comparative genomic studies show that in both bacteria and archaea, most RM system–associated MTases exhibit a patchy phylogenetic distribution indicative of HGT, while orphan MTases on the whole exhibit greater conservation (17, 38, 89, 91, 107). Some orphans appear to have originated from degraded RM systems, and these show the same patchy distribution as intact RM systems (107). Findings from studies using different metrics disagree on the percentage of such RM-derived orphans, with one report putting it as high as 24% (107) and reports of two other related studies estimating that it is much lower (89, 91). Regardless of the percentage, it is clear that a majority of orphans are not recently derived from RM systems and are instead acquired by HGT in their orphaned state (89, 91). It is these true orphans that are the primary contributors to the pattern of greater conservation within taxonomic groups, which is indicative of vertical inheritance and positive selection. Such conserved MTases are referred to as persistent (89), and generally speaking, members of a persistent MTase family recognize the same TS and methylate the same base. Not all orphan MTases are persistent, and not all persistent MTases are orphans, but the two categories largely overlap, particularly for Type II MTases (89).

The availability of large numbers of sequenced bacterial genomes and methylomes has greatly improved our understanding of the origins of orphan MTases and identification of persistent examples. In 2012, Seshasayee and coworkers (107) surveyed 1,227 prokaryotic genomes. Using very stringent search criteria, they identified 914 DNA MTases, of which 60% were deemed non-RM-derived orphans. Using separate methodology in 2016, Blow and coworkers (17) examined 230 prokaryotic genomes, and of the Type II MTases they identified, 52% appeared to be orphans. They defined a persistent MTase as occurring in ≥50% of genomes in a taxonomic group and found that 42% of orphans belonged to one of 19 lineage-specific persistent families, with the remainder being singletons (17). The largest survey to date, by Oliveira and Fang (89), identified 26,582 MTases in 5,568 complete bacterial genomes, a far higher average number than in the previous studies. They restricted their study to species with at least 10 complete genomes available and defined a persistent MTase as occurring in ≥80% of genomes of a given species. Despite this more stringent definition, they found that 52% of species represented in the study harbored at least one persistent MTase.

For many years the study of persistent MTases was limited to three well-known cases: Dam, Dcm, and CcrM. However, the studies above and others have revealed that persistent MTases are more widespread than previously appreciated (17, 38, 89, 107). To a large extent, our understanding of specific functional roles has not kept pace with the discovery of these enzymes. A recent opinion piece has called for a community-wide effort to study these enzymes both as biological phenomena and as potential biomedical targets (89), a declaration with which we strongly concur. In the meantime, several trends have emerged that hint of biological function, at least in general terms.

The seminal 2016 study that identified 19 persistent MTase families also identified the methylated TSs for which they are responsible (17), enabling examination of the methylation properties of orphan and persistent MTases. In contrast to methylation by RM systems, methylation by orphan MTases (whether persistent or not) often resulted in a small number of consistently unmethylated sites throughout the genome (17). Such unmethylated sites were enriched in upstream gene-regulatory regions, often evolutionarily conserved, and sometimes found clustered at chromosomal origins of replication. This signature is suggestive of regulatory roles in gene expression and DNA replication, as these unmethylated sites are likely the result of competition

between the MTase and other DNA-binding proteins for key regulatory sites. The fact that different orphan MTases recognizing different TSs, methylating different bases, and occurring in disparate phylogenetic groups have similar methylome properties suggests convergent evolution toward the same general functions.

Concerning the functions of persistent MTases, a great deal is known about a small number of cases, about which much has already been written, while little or nothing is known about the remainder. We mention some examples here and direct the reader to more complete lists that include uncharacterized examples elsewhere (17, 89).

Dam

Dam, which methylates GATC in many *Gammaproteobacteria*, is the best-studied persistent MTase and has one of the widest known taxonomic ranges. It functions as a maintenance MTase, converting hemimethylated sites to fully methylated sites, but is also capable of efficient de novo methylation (126). Two of its important functions rely on the transient hemimethylated state that exists immediately behind the replication fork during DNA replication (73, 126). First, within the mismatch repair system to correct replication errors, it functions in conjunction with the repair complex MutSLH, which assembles at mismatch sites, directing repair to the unmethylated daughter strand. Second, following DnaA-mediated initiation of DNA replication, clusters of hemimethylated Dam sites located at the origin of replication *oriC* and in the *dnaA* promoter region are bound (sequestered) by SeqA, preventing an inappropriate second round of replication until the next cell cycle. It has recently been shown that Dam also prevents aberrant, *oriC*-independent replication that can occur in certain mutant strains (99). Dam methylation also regulates gene expression, both through the cell cycle–dependent methylation state and through competition with DNA-binding proteins at key sites, as reviewed elsewhere (2, 24, 44, 73, 126). Recent studies have identified unmethylated and undermethylated sites that may be of functional importance (17, 94, 103). The gene encoding Dam is essential in *Vibrio*, where it is required for local chromosomal melting to initiate replication, and it is critical for virulence in *Salmonella*, *Haemophilus*, *Yersinia*, and *Vibrio*, where it regulates expression of genes for fimbrial adhesins, those involved in invasion, and others (126).

Phylogenomic analysis shows that *dam* was acquired in the common ancestor of the *Enterobacterales*, *Pasteurellales*, *Vibrionales*, *Aeromonadales*, and *Alteromonadales* after their divergence from the other *Gammaproteobacteria* orders. The retention of this orphan may have been stabilized by coacquisition of *seqA*, *mutH*, and several other functionally linked genes about the same time (73). This provides one possible model for the selection and retention of orphan MTases and their expansion into lineage-specific families.

CcrM

CcrM, which methylates GANTC, occurs widely in the *Alphaproteobacteria* and was likely acquired by the common ancestor of the orders *Rhodospirillales*, *Sphingomonadales*, *Rhizobiales*, *Rhodobacterales*, and *Caulobacterales* (47). It has been most heavily studied in *Caulobacter crescentus*, a model organism for asymmetric cell division, in which it plays a key role in cell cycle regulation (reviewed in 82). Its expression is very tightly controlled, being activated for a brief period at the end of S phase, immediately following chromosomal replication but before cell division. Thus, the nascent copies remain hemimethylated throughout replication and are converted to a fully methylated state at the onset of cell division. Following cell division and prior to the next round of replication, CcrM is destroyed by the Lon protease in swarmer cells and sequestered at the cell pole in replication-competent stalked cells (132).

Unlike Dam, CcrM is not required for replication initiation or mismatch repair but rather acts primarily as a global transcriptional regulator through temporal cell cycle–dependent methylation state changes (47). The CcrM methylation state affects the binding properties of the transcription factor GcrA (36), resulting in the methylation-dependent expression of yet other global regulators including *dnaA* and *ctrA*, both of which regulate replication (47). GcrA activates *ctrA* expression from the P1 promoter when it is hemimethylated but not when it is fully methylated. During replication, the two daughter chromosomes are hemimethylated on opposite strands, and GcrA activates *creS* only in the copy where the coding strand is methylated (79), imparting asymmetry. In total, CcrM and its regulatory partners influence the transcription of at least 200 genes in *C. crescentus*, at least 40 of which are directly affected by CcrM methylation (131).

Dcm

Dcm is an orphan m5C MTase modifying C<u>C</u>WGG that is conserved throughout the *Escherichia* and *Salmonella* genera (although notably absent in the laboratory strain *E. coli* B). Recent evidence shows that Dcm methylation has gene regulatory effects, reducing the expression of ribosomal protein genes *rplC* and *rpsJ* and the drug resistance transporter *sugE* during stationary phase (77, 78). On the whole, loss of *dcm* results in no obvious phenotype, and so despite decades of study, the reasons for its selective retention remain somewhat unclear. Furthermore, it is cotranscribed with *vsr*, which encodes an endonuclease that specifically repairs the T:G mismatches that result from deamination of m5C, an arrangement likely intended to reduce the mutagenic effects of Dcm methylation (112). Despite this, CCWGG sites show an eightfold increase in mutation rate in natural bacterial populations where *dcm* is present (27), so whatever function Dcm is performing, its hosts appear willing to make this trade-off.

CamA

The orphan MTase designated CamA, which methylates the asymmetric sequence CAAAA<u>A</u>, is ubiquitous in, yet primarily restricted to, the species *Clostridioides difficile* (90). CamA methylation appears to be important for efficient sporulation, particularly for progressing beyond asymmetric division. CamA methylation appears to affect σ^F activation and subsequent expression of *spoIIQ* and *spoIVA* within the sporulation transcriptional cascade. In addition, *camA* mutant cells show overexpression of cell wall–remodeling enzymes and an increase in biofilm production (90). CamA TSs are overrepresented in the regulatory regions upstream of genes involved in sporulation, motility, and membrane transport and significantly overlap with the binding sites of known transcription factors, suggesting that competition for these sites creates regulatory ON/OFF switches (90).

Burkholderia MTases

Two MTases belong to the core genomes of four *Burkholderia* species (*B. cenocepacia*, *B. mallei*, *B. pseudomallei*, and *B. thailandensis*), indicating very strong conservation (89). They have yet to be assigned mnemonics but are referred to here by their REBASE designations from *B. cenocepacia* J2315 (**http://rebase.neb.com**). M.BceJIV is a Type II orphan recognizing GTWW<u>A</u>C, while M.BceJI, recognizing CAC<u>A</u>G, belongs to a complete Type III RM system. Deletion of *bceJIVM* resulted in changes in biofilm morphology and increased aggregation and pellicle formation, while *bceJIM* mutants showed decreased motility (121). Genes differentially expressed between wild-type and mutant strains have been identified, and several of these have relevant TSs in their promoter regions, suggesting direct effects of methylation state on expression (121). However, the

precise effects on the transcriptional program leading to the observed phenotypes remain to be elucidated.

CtsM

An orphan MTase modifying RAATTY is highly conserved in *Campylobacter*, a genus of *Epsilonproteobacteria* that is naturally competent (13, 84, 89, 129). In the *C. jejuni* 81–176 genome, this TS occurs fourfold more often than expected by chance, and the presence of *ctsM* coupled with this excess guarantees that a large proportion of DNA fragments originating in *Campylobacter* cells will bear at least one methylated RAATTY site (13). This methylation mark promotes DNA uptake and is required for efficient natural transformation, thereby establishing preferential HGT corridors within the *Campylobacter* genus (13). This discrimination does not rely on an REase, and further study is needed to determine its mechanism. Methylation at RAATTY sites is also highly conserved in *Acinetobacter* (89) and *Spirochaeta*, and in the latter at least, the presence of unmethylated sites suggests a transcriptional regulatory role (17). Thus, this same epigenetic mark may have been adapted to serve different functions in different lineages.

Other Examples

Other examples not described here include the orphan M.EcoKII (encoded by *yhdJ*) of *E. coli* and *Salmonella* (6, 19, 95); the orphan m5C MTase VchM of *Vibrio cholerae* (25); the MTase of the Type I RM system SpyMEW123I and its orthologs in *Streptococcus pyogenes* (87); the MTase of the Type III RM system StyLTI and its orthologs in *Salmonella* (95); several MTases in cyanobacteria (50); and the m5C MTase M.Hpy99III and its orthologs in *H. pylori*, which have a cognate REase in 13% of observed cases (32). In addition to the examples above, many other recently identified persistent MTases await study. While many persistent MTases likely act through advantageous alterations to gene expression, the examples of Dam and CtsM demonstrate that other functional paradigms are possible.

CONCLUSIONS AND FUTURE DIRECTIONS

At both ends of the change rate spectrum, phase variability and persistence, methylation acts as an epigenetic mark to effect changes in the cell's gene expression program. Any DNA MTase, when introduced into a naive cell, will likely alter gene expression due to the fortuitous juxtaposition of methylation TSs and regulatory-protein-binding sites. If the resulting competition proves advantageous to the cell, both the TS and the MTase will persist. With the MTase established, additional TSs of selective advantage may emerge. The collection of genomes and methylomes sequenced to date is simply a snapshot of dynamic evolution, and we observe MTases in all phases of evolution. Today's orphan singleton may become tomorrow's persistent MTase family, and today's mobile RM system may become tomorrow's established phase-variable or cassette-variable system.

Recent work has shown a number of examples of MTases that may be emergent regulatory agents. For example, DnmA, an orphan MTase modifying the asymmetric sequence GACGAG, has recently been identified in certain *B. subtilis* isolates (88). Although not yet prevalent enough to be deemed persistent (89), it nonetheless exhibits significant effects on gene regulation through competition with the transcription factor ScoC (88). In addition, methylation from HGT-acquired RM systems has significant and immediate impacts on gene expression (72, 127). These examples, coupled with the identification of cassette-variable methylation systems and the finding that orphan singletons exhibit consistently unmethylated sites, suggest that novel methylation patterns can very quickly integrate into a cell's transcriptional program. The fact that persistent MTase

families often adhere to species and higher-order taxonomic boundaries suggests that these programmatic changes can have consequential effects on bacterial evolution.

We currently find ourselves in a very exciting time in the field of bacterial epigenetics. Thanks to the development of SMRT (single-molecule, real-time) sequencing, which can detect methylated sites at single-base resolution (37), the characterization of bacterial methylomes has become commonplace. Over the last decade, we have accumulated enough methylome data to now observe bacterial epigenetic patterns and identify key MTase players on a large scale. It is no coincidence that many of the studies related to novel persistent MTases and their functions cited here are from the last two years. A positive development is that research is beginning to focus on previously understudied groups, such as gram-positive bacteria and archaea, which are as epigenetically rich as gram-negative bacteria. Also intriguing is the emergence of a possible link between methylation and another epigenetic mechanism, phosphorothioation (26). We expect these trends to continue as yet more methylomes are sequenced and existing data in REBASE and other sources are mined. We await the exciting discoveries to come.

DISCLOSURE STATEMENT

The authors work for New England Biolabs, a company that sells research reagents, including restriction enzymes and DNA methyltransferases, to the scientific community.

ACKNOWLEDGMENTS

This article is dedicated to the memory of Donald G. Comb, who made our work possible. The authors are grateful to David Dryden and Elisabeth Raleigh for critical reading of the manuscript and to Lana Saleh for helpful comments.

LITERATURE CITED

1. Adamczyk-Poplawska M, Lower M, Piekarowicz A. 2011. Deletion of one nucleotide within the homonucleotide tract present in the *hsdS* gene alters the DNA sequence specificity of type I restriction-modification system NgoAV. *J. Bacteriol.* 193:6750–59
2. Adhikari S, Curtis PD. 2016. DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.* 40:575–91
3. Alderman MH 3rd, Xiao AZ. 2019. N^6-Methyladenine in eukaryotes. *Cell Mol. Life Sci.* 76:2957–66
4. Anjum A, Brathwaite KJ, Aidley J, Connerton PL, Cummings NJ, et al. 2016. Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in *Campylobacter jejuni* strain NCTC11168. *Nucleic Acids Res.* 44:4581–94
5. Anton BP, Heiter DF, Benner JS, Hess EJ, Greenough L, et al. 1997. Cloning and characterization of the *Bgl*II restriction-modification system reveals a possible evolutionary footprint. *Gene* 187:19–27
6. Anton BP, Mongodin EF, Agrawal S, Fomenkov A, Byrd DR, et al. 2015. Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain of *Escherichia coli* K-12. *PLOS ONE* 10:e0127446
7. Atack JM, Guo C, Litfin T, Yang L, Blackall PJ, et al. 2020. Systematic analysis of REBASE identifies numerous Type I restriction-modification systems with duplicated, distinct *hsdS* specificity genes that can switch system specificity by recombination. *mSystems* 5(4):e00497-20
8. Atack JM, Guo C, Yang L, Zhou Y, Jennings MP. 2020. DNA sequence repeats identify numerous Type I restriction-modification systems that are potential epigenetic regulators controlling phase-variable regulons; phasevarions. *FASEB J.* 34:1038–51
9. Atack JM, Srikhanta YN, Fox KL, Jurcisek JA, Brockman KL, et al. 2015. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nat. Commun.* 6:7828
10. Atack JM, Tan A, Bakaletz LO, Jennings MP, Seib KL. 2018. Phasevarions of bacterial pathogens: Methylomics sheds new light on old enemies. *Trends Microbiol.* 26:715–26

11. Atack JM, Yang Y, Seib KL, Zhou Y, Jennings MP. 2018. A survey of Type III restriction-modification systems reveals numerous, novel epigenetic regulators controlling phase-variable regulons: phasevarions. *Nucleic Acids Res.* 46:3532–42
12. Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD. 1986. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. *Cancer Res.* 46:2917–22
13. Beauchamp JM, Leveque RM, Dawid S, DiRita VJ. 2017. Methylation-dependent DNA discrimination in natural transformation of *Campylobacter jejuni*. *PNAS* 114:E8053–61
14. Ben-Assa N, Coyne MJ, Fomenkov A, Livny J, Robins WP, et al. 2020. Analysis of a phase-variable restriction modification system of the human gut symbiont *Bacteroides fragilis*. *Nucleic Acids Res.* 48(19):11040–53
15. Betlach M, Hershfield V, Chow L, Brown W, Goodman H, Boyer HW. 1976. A restriction endonuclease analysis of the bacterial plasmid controlling the ecoRI restriction and modification of DNA. *Fed. Proc.* 35:2037–43
16. Blakeway LV, Tan A, Lappan R, Ariff A, Pickering JL, et al. 2018. *Moraxella catarrhalis* restriction-modification systems are associated with phylogenetic lineage and disease. *Genome Biol. Evol.* 10:2932–46
17. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, et al. 2016. The epigenomic landscape of prokaryotes. *PLOS Genet*. 12:e1005854
18. Bower EKM, Cooper LP, Roberts GA, White JH, Luyten Y, et al. 2018. A model for the evolution of prokaryotic DNA restriction-modification systems based upon the structural malleability of Type I restriction-modification enzymes. *Nucleic Acids Res.* 46:9067–80
19. Broadbent SE, Balbontin R, Casadesus J, Marinus MG, van der Woude M. 2007. YhdJ, a nonessential CcrM-like DNA methyltransferase of *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* 189:4325–27
20. Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D, et al. 2011. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *PNAS* 108:4494–99
21. Butterer A, Pernstich C, Smith RM, Sobott F, Szczelkun MD, Toth J. 2014. Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res*. 42:5139–50
22. Calisto BM, Pich OQ, Pinol J, Fita I, Querol E, Carpena X. 2005. Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* 351:749–62
23. Callahan SJ, Luyten YA, Gupta YK, Wilson GG, Roberts RJ, et al. 2016. Structure of Type IIL restriction-modification enzyme MmeI in complex with DNA has implications for engineering new specificities. *PLOS Biol*. 14:e1002442
24. Casadesus J. 2016. Bacterial DNA methylation and methylomes. *Adv. Exp. Med. Biol.* 945:35–61
25. Chao MC, Zhu S, Kimura S, Davis BM, Schadt EE, et al. 2015. A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen [corrected]. *PLOS Genet*. 11:e1005666. Correction. 2015. *PLOS Genet.* 11:e1005739
26. Chen C, Wang L, Chen S, Wu X, Gu M, et al. 2017. Convergence of DNA methylation and phosphorothioation epigenetics in bacterial genomes. *PNAS* 114:4501–6
27. Cherry JL. 2018. Methylation-induced hypermutation in natural populations of bacteria. *J. Bacteriol.* 200:e00371-18
28. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, et al. 2006. Multireplicon genome architecture of *Lactobacillus salivarius*. *PNAS* 103:6718–23
29. De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, et al. 2017. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. *FEMS Microbiol. Rev.* 41(Suppl. 1):S3–15
30. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, et al. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359(6379):eaar4120
31. Dunn DB, Smith JD. 1955. Occurrence of a new base in the deoxyribonucleic acid of a strain of *Bacterium coli*. *Nature* 175:336–37
32. Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S. 2019. The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. *Nucleic Acids Res*. 47:2336–48

33. Fagerlund A, Langsrud S, Schirmer BC, Moretro T, Heir E. 2016. Genome analysis of *Listeria monocytogenes* sequence type 8 strains persisting in salmon and poultry processing environments and comparison with related strains. *PLOS ONE* 11:e0151117
34. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, et al. 2012. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* 30:1232–39
35. Feinberg AP, Vogelstein B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89–92
36. Fioravanti A, Fumeaux C, Mohapatra SS, Bompard C, Brilli M, et al. 2013. DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other *Alphaproteobacteria*. *PLOS Genet*. 9:e1003541
37. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, et al. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7:461–65
38. Fullmer MS, Ouellette M, Louyakis AS, Papke RT, Gogarten JP. 2019. The patchy distribution of restriction–modification system genes and the conservation of orphan methyltransferases in Halobacteria. *Genes* 10(3):233
39. Furuta Y, Abe K, Kobayashi I. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res*. 38:2428–43
40. Furuta Y, Kawai M, Uchiyama I, Kobayashi I. 2011. Domain movement within a gene: a novel evolutionary mechanism for protein diversification. *PLOS ONE* 6:e18819
41. Furuta Y, Kobayashi I. 2012. Movement of DNA sequence recognition domains between non-orthologous proteins. *Nucleic Acids Res*. 40:9218–32
42. Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, et al. 1983. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res*. 11:6883–94
43. Gawthorne JA, Beatson SA, Srikhanta YN, Fox KL, Jennings MP. 2012. Origin of the diversity in DNA recognition domains in phasevarion associated *modA* genes of pathogenic *Neisseria* and *Haemophilus influenzae*. *PLOS ONE* 7:e32337
44. Giacomodonato MN, Sarnacki SH, Llana MN, Cerquetti MC. 2009. Dam and its role in pathogenicity of *Salmonella enterica*. *J. Infect. Dev. Ctries.* 3:484–90
45. Gold M, Hurwitz J, Anders M. 1963. The enzymatic methylation of RNA and DNA. *Biochem. Biophys. Res. Commun.* 11:107–14
46. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, et al. 2015. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* 34:169–83
47. Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J. 2014. The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach. *Nucleic Acids Res*. 42:3720–35
48. Gordeeva J, Morozova N, Sierro N, Isaev A, Sinkunas T, et al. 2019. BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res*. 47:253–65
49. Gupta YK, Chan SH, Xu SY, Aggarwal AK. 2015. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat. Commun.* 6:7363
50. Hagemann M, Gartner K, Scharnagl M, Bolay P, Lott SC, et al. 2018. Identification of the DNA methyltransferases establishing the methylome of the cyanobacterium *Synechocystis* sp. PCC 6803. *DNA Res.* 25:343–52
51. Harris AJ, Goldman AD. 2020. The complex phylogenetic relationships of a 4mC/6mA DNA methyltransferase in prokaryotes. *Mol. Phylogenet. Evol.* 149:106837
52. He YF, Li BZ, Li Z, Liu P, Wang Y, et al. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303–7
53. Heitman J. 1993. On the origins, structures and functions of restriction-modification enzymes. *Genet. Eng.* 15:57–108
54. Huang X, Wang J, Li J, Liu Y, Liu X, et al. 2020. Prevalence of phase variable epigenetic invertons among host-associated bacteria. *Nucleic Acids Res.* 48(20):11468–85
55. Humbelin M, Suri B, Rao DN, Hornby DP, Eberle H, et al. 1988. Type III DNA *res*triction and *mod*ification systems *Eco*P1 and *Eco*P15: nucleotide sequence of the *Eco*P1 operon, the *Eco*P15 mod gene and some *Eco*P1 *mod* mutants. *J. Mol. Biol.* 200:23–29

56. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–3
57. Iyer LM, Tahiliani M, Rao A, Aravind L. 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8:1698–710
58. Iyer LM, Zhang D, Burroughs AM, Aravind L. 2013. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res*. 41:7635–55
59. Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V. 1983. Cytosine modification in DNA by *Bcn*I methylase yields N^4-methylcytosine. *FEBS Lett*. 161:131–34
60. Jeltsch A, Pingoud A. 1996. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.* 42:91–96
61. Jen FE-C, Scott AL, Tan A, Seib KL, Jennings MP. 2020. Random switching of the ModA11 Type III DNA methyltransferase of *Neisseria meningitidis* regulates Entner–Doudoroff aldolase expression by a methylation change in the *eda* promoter region. *J. Mol. Biol.* 432(21):5835–42
62. Jin B, Li Y, Robertson KD. 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* 2:607–17
63. Johnson TB, Coghill RD. 1925. Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle *Bacillus*. *J. Am. Chem. Soc.* 47:2838–44
64. Jonsson AB, Nyberg G, Normark S. 1991. Phase variation of gonococcal pili by frameshift mutation in *pilC*, a novel gene for pilus assembly. *EMBO J.* 10:477–88
65. Jurkowska RZ, Jeltsch A. 2016. Mechanisms and biological roles of DNA methyltransferases and DNA methylation: from past achievements to future challenges. *Adv. Exp. Med. Biol.* 945:1–17
66. Kita K, Tsuda J, Kato T, Okamoto K, Yanase H, Tanaka M. 1999. Evidence of horizontal transfer of the *Eco*O109I restriction-modification gene to *Escherichia coli* chromosomal DNA. *J. Bacteriol.* 181:6822–27
67. Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res*. 29:3742–56
68. Krebes J, Morgan RD, Bunk B, Sproer C, Luong K, et al. 2014. The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res*. 42:2415–32
69. Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324:929–30
70. Kwun MJ, Oggioni MR, De Ste Croix M, Bentley SD, Croucher NJ. 2018. Excision-reintegration at a pneumococcal phase-variable restriction-modification locus drives within- and between-strain epigenetic differentiation and inhibits gene acquisition. *Nucleic Acids Res*. 46:11438–53
71. Lin LF, Posfai J, Roberts RJ, Kong H. 2001. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *PNAS* 98:2740–45
72. Liu G, Jiang YM, Liu YC, Han LL, Feng H. 2020. A novel DNA methylation motif identified in *Bacillus pumilus* BA06 and possible roles in the regulation of gene expression. *Appl. Microbiol. Biotechnol.* 104:3445–57
73. Løbner-Olesen A, Skovgaard O, Marinus MG. 2005. Dam methylation: coordinating cellular processes. *Curr. Opin. Microbiol.* 8:154–60
74. Malone T, Blumenthal RM, Cheng X. 1995. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* 253:618–32
75. Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, et al. 2014. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat. Commun.* 5:5055
76. Martin JL, McMillan FM. 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* 12:783–93
77. Militello KT, Mandarano AH, Varechtchouk O, Simon RD. 2014. Cytosine DNA methylation influences drug resistance in *Escherichia coli* through increased *sugE* expression. *FEMS Microbiol. Lett*. 350:100–6
78. Militello KT, Simon RD, Qureshi M, Maines R, VanHorne ML, et al. 2012. Conservation of Dcm-mediated cytosine DNA methylation in *Escherichia coli*. *FEMS Microbiol. Lett.* 328:78–85

79. Mohapatra SS, Fioravanti A, Vandame P, Spriet C, Pini F, et al. 2020. Methylation-dependent transcriptional regulation of crescentin gene (*creS*) by GcrA in *Caulobacter crescentus*. *Mol. Microbiol.* 114:127–39
80. Morgan RD, Dwinell EA, Bhatia TK, Lang EM, Luyten YA. 2009. The MmeI family: type II restriction-modification enzymes that employ single-strand modification for host protection. *Nucleic Acids Res.* 37:5208–21
81. Morgan RD, Luyten YA. 2009. Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity. *Nucleic Acids Res.* 37:5222–33
82. Mouammine A, Collier J. 2018. The impact of DNA methylation in Alphaproteobacteria. *Mol. Microbiol.* 110:1–10
83. Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* 40:307–33
84. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, et al. 2012. The methylomes of six bacteria. *Nucleic Acids Res.* 40:11450–62
85. Murray NE. 2002. Immigration control of DNA in bacteria: self versus non-self. *Microbiology* 148:3–20
86. Naito T, Kusano K, Kobayashi I. 1995. Selfish behavior of restriction-modification systems. *Science* 267:897–99
87. Nye TM, Jacob KM, Holley EK, Nevarez JM, Dawid S, et al. 2019. DNA methylation from a Type I restriction modification system influences gene expression and virulence in *Streptococcus pyogenes*. *PLOS Pathog.* 15:e1007841
88. Nye TM, van Gijtenbeek LA, Stevens AG, Schroeder JW, Randall JR, et al. 2020. Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in *Bacillus subtilis*. *Nucleic Acids Res.* 48:5332–48
89. Oliveira PH, Fang G. 2021. Conserved DNA methyltransferases: a window into fundamental mechanisms of epigenetic regulation in bacteria. *Trends Microbiol.* 29(1):28–40
90. Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, et al. 2020. Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nat. Microbiol.* 5:166–80
91. Oliveira PH, Touchon M, Rocha EP. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 42:10618–31
92. Oliveira PH, Touchon M, Rocha EP. 2016. Regulation of genetic flux between bacteria by restriction-modification systems. *PNAS* 113:5658–63
93. Orlowski J, Bujnicki JM. 2008. Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.* 36:3552–69
94. Payelleville A, Legrand L, Ogier JC, Roques C, Roulet A, et al. 2018. The complete methylome of an entomopathogenic bacterium reveals the existence of loci with unmethylated adenines. *Sci. Rep.* 8:12091
95. Pirone-Davies C, Hoffmann M, Roberts RJ, Muruvanda T, Timme RE, et al. 2015. Genome-wide methylation patterns in *Salmonella enterica* subsp. *enterica* serovars. *PLOS ONE* 10:e0123639
96. Posfai J, Bhagwat AS, Posfai G, Roberts RJ. 1989. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* 17:2421–35
97. Radlinska M, Bujnicki JM. 2001. Cloning of enterohemorrhagic *Escherichia coli* phage VT-2 dam methyltransferase. *Acta Microbiol. Pol.* 50:161–67
98. Radlinska M, Bujnicki JM, Piekarowicz A. 1999. Structural characterization of two tandemly arranged DNA methyltransferase genes from *Neisseria gonorrhoeae* MS11: N4-cytosine specific M.*Ngo*MXV and nonfunctional 5-cytosine-type M.*Ngo*Morf2P. *Proteins* 37:717–28
99. Raghunathan N, Goswami S, Leela JK, Pandiyan A, Gowrishankar J. 2019. A new role for *Escherichia coli* Dam DNA methylase in prevention of aberrant chromosomal replication. *Nucleic Acids Res.* 47:5698–711
100. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, et al. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31:1805–12
101. Rochepeau P, Selinger LB, Hynes MF. 1997. Transposon-like structure of a new plasmid-encoded restriction-modification system in *Rhizobium leguminosarum* VF39SM. *Mol. Gen. Genet.* 256:387–96
102. Sanchez-Buso L, Golparian D, Parkhill J, Unemo M, Harris SR. 2019. Genetic variation regulates the activation and specificity of Restriction-Modification systems in *Neisseria gonorrhoeae*. *Sci. Rep.* 9:14685

103. Sánchez-Romero MA, Olivenza DR, Gutiérrez G, Casadesús J. 2020. Contribution of DNA adenine methylation to gene expression heterogeneity in *Salmonella enterica*. *Nucleic Acids Res*. 48(21):11857–67
104. Schluckebier G, O'Gara M, Saenger W, Cheng X. 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* 247:16–20
105. Seib KL, Pigozzi E, Muzzi A, Gawthorne JA, Delany I, et al. 2011. A novel epigenetic regulator associated with the hypervirulent *Neisseria meningitidis* clonal complex 41/44. *FASEB J*. 25:3622–33
106. Seib KL, Srikhanta YN, Atack JM, Jennings MP. 2020. Epigenetic regulation of virulence and immunoevasion by phase-variable restriction-modification systems in bacterial pathogens. *Annu. Rev. Microbiol.* 74:655–71
107. Seshasayee AS, Singh P, Krishna S. 2012. Context-dependent conservation of DNA methyltransferases in bacteria. *Nucleic Acids Res*. 40:7066–73
108. Sibley MH, Raleigh EA. 2004. Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives. *Nucleic Acids Res*. 32:522–34
109. Sitaraman R, Denison AM, Dybvig K. 2002. A unique, bifunctional site-specific DNA recombinase from *Mycoplasma pulmonis*. *Mol. Microbiol.* 46:1033–40
110. Skoglund A, Bjorkholm B, Nilsson C, Andersson AF, Jernberg C, et al. 2007. Functional analysis of the M.HpyAIV DNA methyltransferase of *Helicobacter pylori*. *J. Bacteriol.* 189:8914–21
111. Skowron PM, Anton BP, Czajkowska E, Zebrowska J, Sulecka E, et al. 2017. The third restriction-modification system from *Thermus aquaticus* YT-1: solving the riddle of two TaqII specificities. *Nucleic Acids Res*. 45:9005–18
112. Sohail A, Lieb M, Dar M, Bhagwat AS. 1990. A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. *J. Bacteriol.* 172:4214–21
113. Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, et al. 2009. Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. *PLOS Pathog.* 5:e1000400
114. Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, Kwok T, et al. 2011. Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. *PLOS ONE* 6:e27569
115. Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP. 2005. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *PNAS* 102:5547–51
116. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–35
117. Tan A, Hill DM, Harrison OB, Srikhanta YN, Jennings MP, et al. 2016. Distribution of the type III DNA methyltransferases modA, modB and modD among *Neisseria meningitidis* genotypes: implications for gene regulation and virulence. *Sci. Rep.* 6:21015
118. Tauseef I, Harrison OB, Wooldridge KG, Feavers IM, Neal KR, et al. 2011. Influence of the combination and phase variation status of the haemoglobin receptors HmbR and HpuAB on meningococcal virulence. *Microbiology* 157:1446–56
119. Tyndall C, Lehnherr H, Sandmeier U, Kulik E, Bickle TA. 1997. The type IC *hsd* loci of the enterobacteria are flanked by DNA with high homology to the phage P1 genome: implications for the evolution and spread of DNA restriction systems. *Mol. Microbiol.* 23:729–36
120. Vale FF, Megraud F, Vitor JM. 2009. Geographic distribution of methyltransferases of *Helicobacter pylori*: evidence of human host population isolation and migration. *BMC Microbiol*. 9:193
121. Vandenbussche I, Sass A, Pinto-Carbo M, Mannweiler O, Eberl L, Coenye T. 2020. DNA methylation epigenetically regulates gene expression in *Burkholderia cenocepacia* and controls biofilm formation, cell aggregation, and motility. *mSphere* 5(4):e00455–20
122. Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77:53–72
123. Weiser JN, Maskell DJ, Butler PD, Lindberg AA, Moxon ER. 1990. Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J. Bacteriol.* 172:3304–9
124. Willemse N, Schultsz C. 2016. Distribution of Type I restriction-modification systems in *Streptococcus suis*: an outlook. *Pathogens* 5(4):62
125. Wilson GG, Murray NE. 1991. Restriction and modification systems. *Annu. Rev. Genet.* 25:585–627
126. Wion D, Casadesus J. 2006. N^6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4:183–92

127. Wu D, Wang Y, Xu X. 2020. Effects of a type I RM system on gene expression and glycogen catabolism in *Synechocystis* sp. PCC 6803. *Front. Microbiol.* 11:1258
128. Yamaichi Y, Chao MC, Sasabe J, Clark L, Davis BM, et al. 2015. High-resolution genetic analysis of the requirements for horizontal transmission of the ESBL plasmid from *Escherichia coli* O104:H4. *Nucleic Acids Res*. 43:348–60
129. Zautner AE, Goldschmidt AM, Thurmer A, Schuldes J, Bader O, et al. 2015. SMRT sequencing of the *Campylobacter coli* BfR-CA-9557 genome sequence reveals unique methylation motifs. *BMC Genom*. 16:1088
130. Zhao L, Song Y, Li L, Gan N, Brand JJ, Song L. 2018. The highly heterogeneous methylated genomes and diverse restriction-modification systems of bloom-forming *Microcystis*. *Harmful Algae* 75:87–93
131. Zhou B, Schrader JM, Kalogeraki VS, Abeliuk E, Dinh CB, et al. 2015. The global regulatory architecture of transcription during the *Caulobacter* cell cycle. *PLOS Genet*. 11:e1004831
132. Zhou X, Wang J, Herrmann J, Moerner WE, Shapiro L. 2019. Asymmetric division yields progeny cells with distinct modes of regulating cell cycle-dependent chromosome methylation. *PNAS* 116:15661–70
133. Ziebuhr W, Krimmer V, Rachid S, Lossner I, Gotz F, Hacker J. 1999. A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Mol. Microbiol.* 32:345–56