

# Deciphering bacterial epigenomes using modern sequencing technologies

John Beaulaurier, Eric E. Schadt<sup>1</sup> and Gang Fang<sup>1</sup>\*

**Abstract** | Prokaryotic DNA contains three types of methylation: N6-methyladenine, N4-methylcytosine and 5-methylcytosine. The lack of tools to analyse the frequency and distribution of methylated residues in bacterial genomes has prevented a full understanding of their functions. Now, advances in DNA sequencing technology, including single-molecule, real-time sequencing and nanopore-based sequencing, have provided new opportunities for systematic detection of all three forms of methylated DNA at a genome-wide scale and offer unprecedented opportunities for achieving a more complete understanding of bacterial epigenomes. Indeed, as the number of mapped bacterial methylomes approaches 2,000, increasing evidence supports roles for methylation in regulation of gene expression, virulence and pathogen–host interactions.

## Phase variation

A means by which reversible variation of protein expression is achieved in bacteria, often in an ON/OFF manner. The process creates phenotypic diversity in a clonally expanded population and allows the colony to survive in rapidly changing environments.

## Adaptive selection

An evolutionary process through which surviving organisms accumulate genetic changes that lead to a fitness advantage over their progenitors.

Department of Genetics and Genomic Sciences and Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

\*e-mail: gang.fang@mssm.edu

<https://doi.org/10.1038/s41576-018-0081-3>

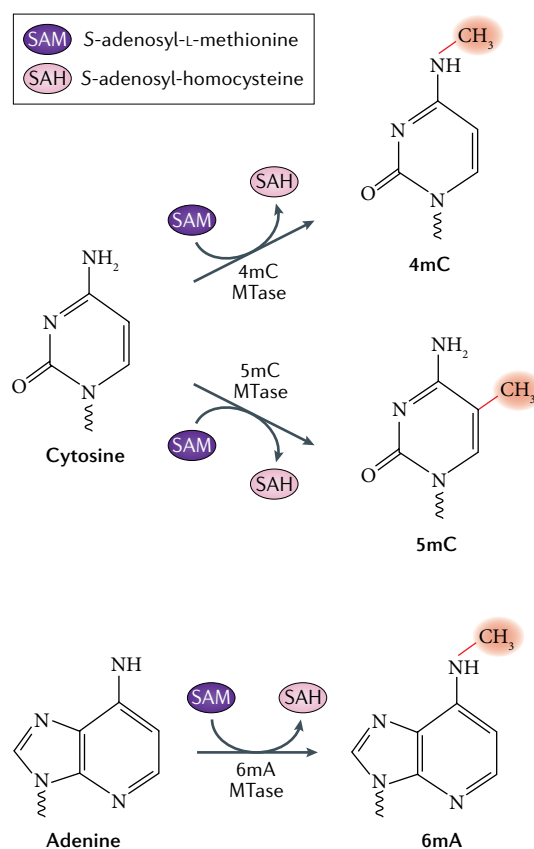
DNA methylation was discovered in bacteria more than a half century ago<sup>1</sup>. It is now known that modification of the four canonical DNA bases by methylation can act as an epigenetic regulator — that is, it can impart distinct and reversible regulatory states to identical genetic sequences. In eukaryotes, epigenetic regulation can occur at multiple levels: DNA methylation, nucleosome positioning, histone variants and histone modifications. By contrast, bacteria lack histones and nucleosomes; therefore, DNA methylation is their primary means of epigenetic gene regulation.

Three different forms of DNA methylation exist in bacterial genomes: N6-methyladenine (6mA), which is the most prevalent form; N4-methylcytosine (4mC); and 5-methylcytosine (5mC). Although 5mC is the dominant form in eukaryotes, 6mA is the most prevalent form in prokaryotes. DNA is methylated by methyltransferase (MTase) enzymes, which transfer a methyl group from S-adenosyl-L-methionine (SAM) to the appropriate position on target bases (FIG. 1). Importantly, only a select few sequence motifs in each bacterial genome are targeted by MTases; for example, in *Escherichia coli*, 5'-GATC-3' is targeted by DNA adenine methylase (Dam) and 5'-CCWGG-3' by DNA-cytosine methyltransferase (Dcm). However, nearly every occurrence of the target motifs is methylated<sup>2</sup>. The MTase specificity domain that determines the target motif varies widely across species, resulting in a large diversity of methylated motifs across the bacterial kingdom.

MTases function either alongside a cognate restriction enzyme as part of a restriction–modification (RM) system or as ‘orphans’ that lack a cognate restriction enzyme. DNA methylation mediated by both types of

MTases has been found to play important regulatory roles in bacteria<sup>2–12</sup>. RM systems protect cells from invading DNA by methylating endogenous DNA and cleaving non-methylated foreign DNA<sup>3,4</sup>. RM systems are divided into three main categories based on the subunits involved and the precise site of DNA restriction<sup>13–16</sup> (FIG. 2). Orphan MTases, such as Dam in Gammaproteobacteria, are thought to regulate DNA replication and gene expression, among other functions<sup>2</sup>. There is also emerging evidence that heterogeneity in methylation patterns within bacterial populations (often caused by phase variation of MTases) can promote adaptive selection by generating heterogeneity in gene expression and cellular phenotypes beyond those provided by genetic variation alone<sup>17,18</sup>. The vast majority of the >6,000 bacterial genomes sequenced to date have been found to encode MTases and are, therefore, likely to be subject to DNA methylation<sup>19,20</sup>. Nonetheless, the precise sequence targets and biological roles of most MTases remain unknown<sup>21</sup>, largely owing to the historical lack of high-throughput tools for detecting 6mA and 4mC. Indeed, while method development for detecting eukaryotic 5mC has flourished over the past few decades<sup>22–25</sup>, only modest methodological advances for detecting the principle forms of DNA methylation in bacterial genomes have been made over the same period.

The recent introduction of new sequencing technologies is beginning to address this problem (TABLE 1). In particular, single-molecule, real-time (SMRT) sequencing has enabled all three major forms of bacterial DNA methylation to be detected simultaneously for the first time, and this technology has been used



**Fig. 1 | Primary types of DNA methylation in bacteria.** Chemical structures are shown for the most common forms of DNA methylation in bacteria, including N4-methylcytosine (4mC), 5-methylcytosine (5mC) and N6-methyladenine (6mA). In each instance, a methyltransferase (MTase) transfers a methyl group ( $\text{CH}_3$ ) from S-adenosyl-L-methionine to the unmodified nucleotide, producing a methylated nucleotide and S-adenosyl-homocysteine.

to generate most of the > 2,000 mapped bacterial and archaeal methylomes<sup>3,20,26–31</sup> that are currently available in the centralized [REBASE database](#)<sup>19</sup>. These methylomes represent a wide variety of isolates from more than 750 distinct species, including common human pathogens such as *Salmonella enterica* ( $n = 150$ ), *E. coli* ( $n = 123$ ), *Klebsiella pneumoniae* ( $n = 93$ ) and *Staphylococcus aureus* ( $n = 47$ ).

Here, we review the currently available methods for mapping bacterial methylomes, with a focus on cutting-edge technologies such as SMRT sequencing and Oxford Nanopore sequencing<sup>32</sup>. We discuss their potential to provide us with fully characterized methylomes that contain not only the full set of methylated positions and targeted motifs but also a complete map of the MTases and RM systems responsible for each methylated motif. We provide an overview of the insights have afforded and discuss how they might be used in the future to obtain a more complete understanding of bacterial epigenomes and the complex roles they play in defining interactions between bacteria and their host organisms. We do not attempt to review the rich

history and foundations of bacterial epigenetics, which have been thoroughly reviewed elsewhere<sup>2,4,21,33,34</sup>.

### Early methods for mapping methylomes

The bulk of methodological development for DNA methylation detection has historically been devoted to characterizing 5mC in higher eukaryotes, largely because the biological importance of 5mC in mammalian cells has been recognized for more than half a century<sup>35–37</sup>. However, characterization of bacterial methylomes requires alternative methods that can detect the more prevalent 6mA and 4mC in addition to 5mC. A number of different approaches have traditionally been used to characterize DNA methylation in bacterial genomes (TABLE 1).

**Restriction enzyme-based mapping.** Prokaryotic MTases are known to primarily target specific sequence motifs for methylation. The genome-wide methylation status of these motifs can often be deduced by digesting genomic DNA with one or more methyl-sensitive restriction enzymes of known specificities and analysing the pattern of cut and uncut restriction sites by next-generation sequencing (NGS)<sup>38,39</sup>. This approach is robust, reliable and accurate but is limited to the study of methylation motifs that perfectly or partially match the known specificities of available restriction enzymes. Thus, although still useful for assessing methylation events within known sequence motifs, it is generally not suitable for discovering new motifs.

**Sanger sequencing-based mapping.** Theoretically, the most common forms of bacterial DNA methylation can be detected as a by-product of Sanger sequencing because the presence of 4mC, 5mC and 6mA in the DNA template affects the amplitude of peaks in the sequencing trace. Although several studies have used this method to investigate the methylomes of pathogenic bacteria<sup>40–44</sup>, technical limitations, including subtle peak signatures and the low throughput of Sanger sequencing, have prevented it from achieving wider usage<sup>32</sup>.

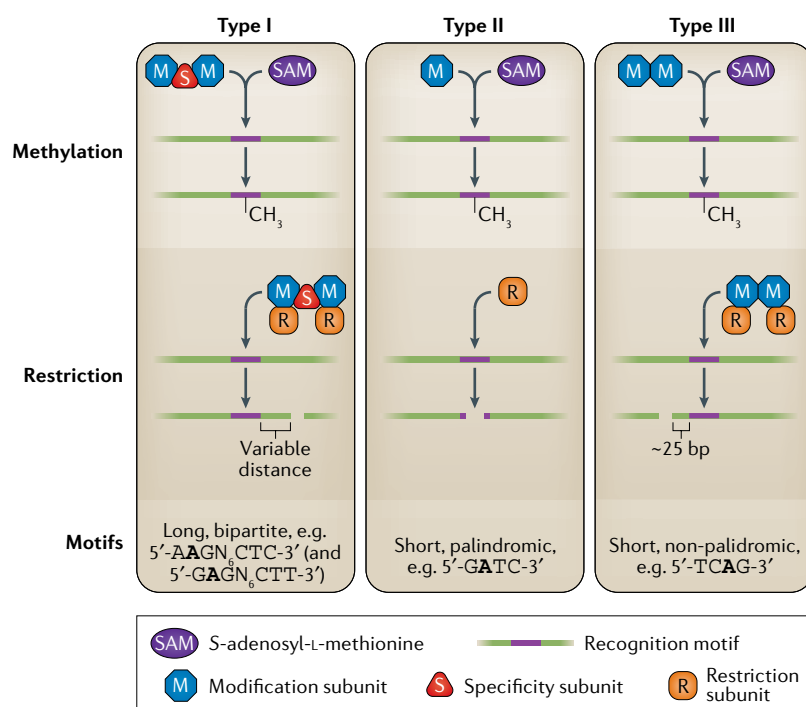
**Bisulfite sequencing-based mapping.** Despite its reputation as the gold standard for characterizing 5mC in eukaryotic genomes (owing to its high sensitivity, accuracy and compatibility with NGS technologies), bisulfite sequencing has only quite recently been applied to the study of 5mC in bacteria<sup>45,46</sup>. More recently, it has been shown that treating genomic DNA with ten-eleven translocation (TET) enzymes before bisulfite treatment makes it possible to characterize both 5mC and 4mC bacterial methylomes<sup>47</sup>. Before bisulfite treatment, TET enzymes oxidize 5mC to 5-carboxylcytosine (5caC), which is subsequently read as a thymine in bisulfite sequencing data. By using a combination of standard bisulfite sequencing and 4mC-TET-assisted bisulfite sequencing (4mC-TAB-seq), it becomes possible to distinguish some fraction of 4mC positions from the 5mC positions; ideally, a sufficient number are detected to permit identification of the 4mC motif<sup>47</sup>. However, 6mA remains undetectable with this approach.

#### Methylomes

The entirety of DNA methylation marks across genomes.

#### Bisulfite sequencing

The treatment of DNA with bisulfite chemically converts unmethylated cytosines to uracils. As methylated cytosines are unaffected, the location of methylation can be identified by sequencing the bisulfite-treated DNA.



**Fig. 2 | Types of restriction–modification systems.** Restriction–modification (RM) systems are divided into three categories based on the subunits involved and the precise site of DNA restriction. Type I systems are composed of a single enzyme containing restriction, modification and specificity subunits, and multisubunit complexes are necessary for both modification and restriction. They target bipartite motifs, that is, two short sub-motifs of a specific sequence that are separated by a fixed number of nonspecific nucleotides. Cleavage can occur up to several kilobases away from the non-methylated motif site<sup>14</sup>. Type II RM systems are composed of distinct methyltransferase (MTase) and restriction enzymes that target short palindromic motifs. The restriction enzyme cleaves DNA close to or within the non-methylated motif sites<sup>15</sup>. Type III RM systems consist of complexes of multiple modification and restriction subunits, with the specificity element contained in the MTase. Short, non-palindromic sequences are targeted for methylation, and non-methylated motif sites are targeted by a separate restriction enzyme, which must bind with the MTase to achieve sequence specificity and cut at a location roughly 25 bp from the motif<sup>16</sup>. In each panel, the nucleotide targeted for methylation within a motif is indicated in bold text.

### Newer methods for mapping methylomes

Recent innovations in so-called third-generation sequencing technologies, including SMRT sequencing and nanopore sequencing (FIG. 3), make it possible to directly interrogate native DNA molecules without PCR amplification. Importantly, these approaches retain chemical modifications in the DNA and enable many of them to be detected for the first time at a genome-wide scale.

**Direct detection of DNA methylation with single-molecule, real-time sequencing.** SMRT sequencing, which is available in the commercialized RS II and Sequel instruments manufactured by Pacific Biosciences, is the first third-generation sequencing technology with a record of successfully characterizing bacterial methylomes. SMRT sequencing can simultaneously report nucleotide sequence and all three major types of DNA methylation in bacteria, albeit at different sensitivities (high for 6mA, moderate for 4mC and low for 5mC) owing to the signal-to-noise ratios specific to each modification type<sup>3,18,20,27,28,48</sup>.

In SMRT sequencing, each template molecule consists of a double-stranded native DNA fragment that has been circularized by ligating hairpin adaptors to each end<sup>26</sup> (FIG. 3a). A DNA polymerase enzyme is bound to the hairpin-adapted template molecule, and a specially designed surface chemistry immobilizes the polymerase–template complex at the base of a zeptolitre-scale observation chamber, called a zero-mode waveguide (ZMW), which limits background fluorescence originating outside this small observation chamber<sup>26,49</sup> (FIG. 3b). Sequencing by synthesis is initiated, and the DNA polymerase proceeds around the circularized DNA template multiple times, generating multiple subreads. During each base incorporation event, a fluorescently labelled deoxyribonucleoside triphosphate (dNTP) complementary to the template base is briefly immobilized in the ZMW observation window by the polymerase. A camera captures the resulting fluorescent pulse, and because each of the four canonical bases has a different fluorescent label, the series of pulses observed as the complementary DNA strand is synthesized can be used to construct the sequence read. Although base calling is accomplished by monitoring the order of dNTP incorporation events, DNA modifications are detected by identifying changes in the kinetics of the polymerase as it translocates along the DNA template (FIG. 3c). Specifically, the polymerase kinetics is described by the inter-pulse duration (IPD), which is the time interval between pulses that indicate nucleotide incorporation events. It has been shown that the IPD can be perturbed by the primary and secondary structures of the DNA template molecules<sup>26</sup> and by covalent DNA modifications, including 4mC, 6mA, 5mC, 5-hydroxymethylcytosine (5hmC) and other types of DNA methylation and damage<sup>27,48,50,51</sup>. The signal-to-noise ratio for 4mC and 6mA is sufficiently high that they can be directly detected in native DNA<sup>3,18,48</sup>. However, detection of 5mC and 5hmC requires either high sequencing depth or additional steps to convert them to 5-formylcytosine and 5caC, both of which have higher signal-to-noise ratios<sup>48,51</sup>.

In order to detect modified nucleotides, IPD values from native DNA sequencing data can be compared with control IPD values from either methylation-free whole-genome amplified (WGA) DNA or precomputed in silico IPD models. The in silico model is trained using large amounts of sequencing data from unmodified DNA and consists of predicted IPD values for a given local sequence context<sup>28</sup>. The local sequence context surrounding the site of nucleotide incorporation strongly affects the processivity of the polymerase, and resulting fluctuations in IPD values must, therefore, be accounted for when looking for IPD deviations caused by a base modification event<sup>27,48</sup>. Owing to the extent of contact between the polymerase and the template DNA molecule, a modified base can affect the IPD values both upstream and downstream of the modified position. The resulting IPD signatures differ between 6mA, 4mC and 5mC, so it is usually possible to assign a methylation type to an observed methylation motif<sup>20,27</sup>. The vast majority of SMRT methylome studies have used a consensus approach to assess IPD values in aligned native reads from data of high sequencing depth.

Table 1 | Methods currently used to detect methylated DNA in prokaryotes at single-nucleotide resolution

Technique	Modifications detected	Advantages	Limitations	Notes	Refs
Bisulfite sequencing	5mC, 5hmC, 4mC	Gold standard for detection of 5mC modifications	<ul style="list-style-type: none"> <li>Bisulfite treatment fragments library DNA molecules</li> <li>4mC detection requires additional conversion steps</li> <li>Cannot detect 6mA</li> </ul>	<ul style="list-style-type: none"> <li>Commonly used for single-nucleotide resolution of 5mC in eukaryotes</li> <li>Application in prokaryotic genomes has been limited</li> </ul>	45,46,165
Restriction enzyme digest followed by next-generation sequencing	4mC, 5mC, 6mA	<ul style="list-style-type: none"> <li>Gold standard for experimental validation of methylated sites</li> <li>Highly sensitive</li> <li>Works well with limited input DNA</li> </ul>	Limited choice of restriction enzymes with known target sequence specificity	Characterization of new restriction–modification systems through new techniques is adding to the pool of available restriction enzymes with known specificities	105
Single-molecule, real-time sequencing	4mC, 5mC, 5hmC, 6mA	<ul style="list-style-type: none"> <li>Long reads enable phasing of multiple methylated positions</li> <li>Reveals both methylated motifs and methylation sites at single-nucleotide resolution and single-molecule resolution</li> </ul>	5mC detection requires additional conversion steps or very deep sequencing	Enabled the mapping of the first complete bacterial epigenome (>2,000 mapped to date)	3,18,27,28,48
Nanopore sequencing	4mC, 5mC, 5hmC, 6mA	Long reads hold promise of phasing multiple methylated positions	Accurate modification detection complicated by noisy signal	Modification detection algorithms under active development	61,64–67

4mC, N4-methylcytosine; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 6mA, N6-methyladenine.

By comparing the native IPD values at a genomic position with either the predicted unmodified IPD value (using the *in silico* model) or a distribution of control IPD values obtained from WGA DNA, a simple statistical test can be used to identify methylated positions where the native and control IPD values diverge<sup>3,27,52</sup>. Alternative methods and statistical models have been proposed for methylation detection in conditions where the sequencing depth is expected to be low. For example, in metagenomic sequencing, low-abundance species would not be expected to be well covered, and for bacterial populations exhibiting heterogeneous methylation, one would not expect to find uniform methylation patterns in a given sample<sup>18,53</sup>.

**Direct detection of DNA methylation with nanopore sequencing.** Nanopore sequencing has been under active development for decades, but recent progress has led to the release in 2014 of MinION, the first commercially available sequencing platform by Oxford Nanopore Technologies<sup>54–61</sup> and, more recently, the release of the GridION and PromethION. In this technology, genetically engineered protein nanopores are placed in a lipid membrane, across which a voltage is applied to drive the negatively charged single-stranded DNA (ssDNA) through the nanopores for sequencing. Multiple protocols are available for constructing libraries for nanopore sequencing, which all involve the ligation of adaptor sequences coupled with a motor protein to double-stranded DNA (dsDNA) fragments (FIG. 3d). The tethering adaptor sequences help to concentrate the DNA fragments near the nanopore-containing lipid membrane, while the motor protein facilitates the processive ratcheting of ssDNA through the protein nanopore at a fixed rate during sequencing. Sensors monitor the current through each nanopore

during this process and detect variations caused by the translocation of the polynucleotide strand obstructing the channel (FIG. 3e). These current fluctuations are a function of the roughly six specific nucleotides occupying the constricted part of the nanopore channel at a given moment (FIG. 3f) and are processed by a proprietary recursive neural network to construct the sequence of nucleotides in the read<sup>61</sup>.

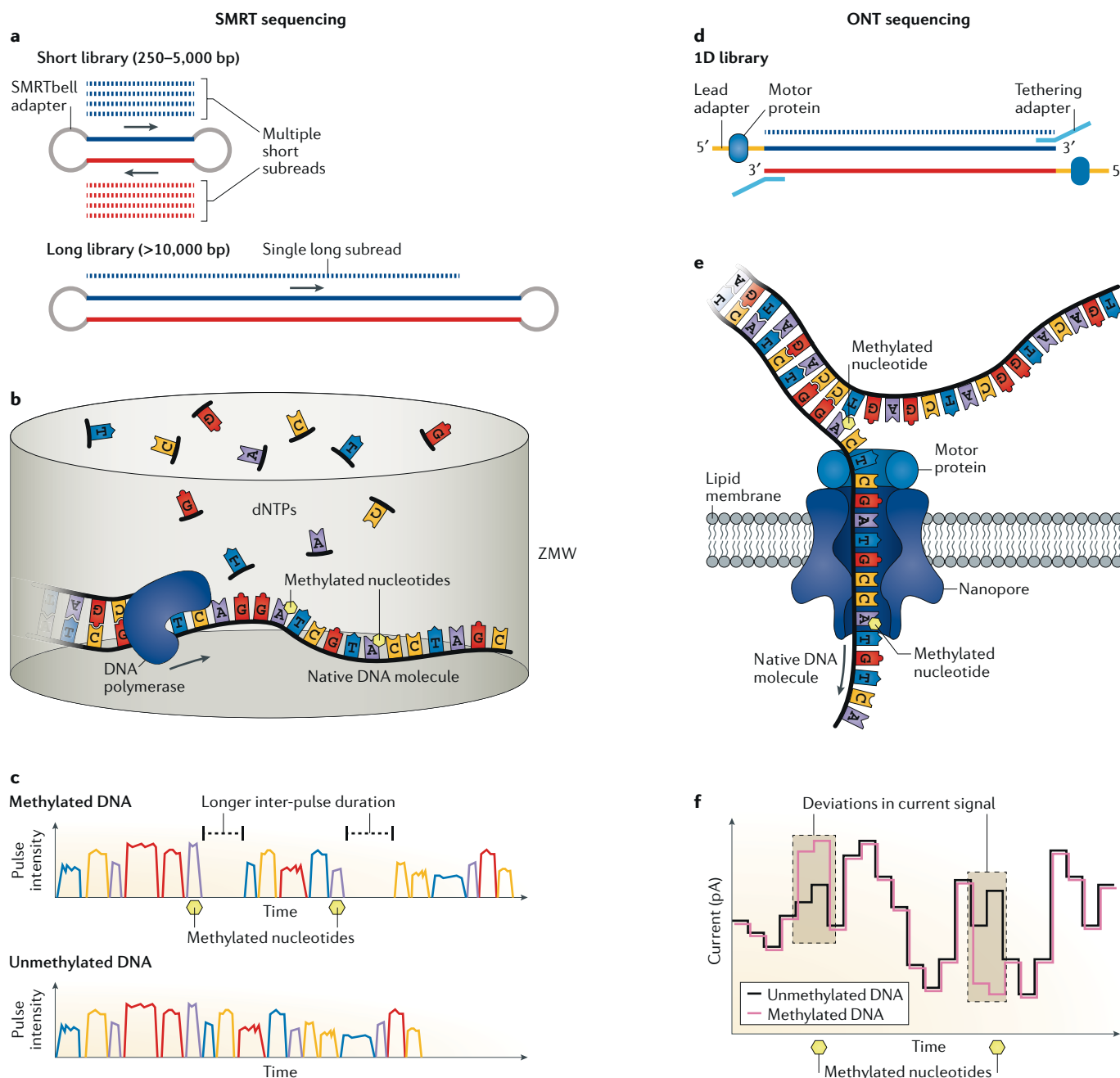
Although the vast majority of research applications to date have focused on using MinION to call the four canonical bases, current signals have been shown to differ between canonical bases and covalently modified nucleotides, which provides the possibility of detecting chemical DNA modifications<sup>62</sup>. However, the presence of modified bases can potentially complicate the base calling process, which relies on characteristic current levels produced as each *k*-mer (a combination of nucleotides of length *k*) passes through the nanopore. The presence of multiple types of base modifications greatly expands the set of possible *k*-mers beyond those constructed exclusively from the four canonical bases, which introduces substantial computational challenges.

Early attempts to detect methylation during nanopore sequencing, using a variety of protein nanopore configurations and experimental conditions, focused on eukaryotic applications and, therefore, were limited to 5mC and 5hmC detection<sup>54,55,61,63</sup>. However, the introduction of the MinION device has recently broadened the development focus towards characterizing prokaryotic methylomes<sup>64–67</sup>. For instance, a variable order hidden Markov model (HMM) was trained to identify methylation events in bacterial genomes<sup>64</sup>. When paired with a hierarchical Dirichlet process (HDP) to learn current distributions from the MinION device, it can detect both 5mC and 6mA at the specific motifs included in the training data. This HMM–HDP approach can

**Hidden Markov model (HMM).** A mathematical concept that describes a finite set of ‘states’ and a probabilistic model for transitioning from one state to another. The probability associated with each transition can be derived from training sets. HMMs are valuable because they enable a search or alignment algorithm to be built on firm probabilistic bases.

**Hierarchical Dirichlet process (HDP).** A non-parametric Bayesian approach for modelling a collection of mixture distributions that share mixture components.





**Fig. 3 | Technologies for detecting DNA methylation through direct sequencing of native DNA molecules.** **a** | Libraries for single-molecule, real-time (SMRT) sequencing from Pacific Biosciences consist of double-stranded DNA fragments flanked by hairpin SMRTbell adaptors that permit the polymerase to process both strands of the template. The libraries can be configured to accommodate the requirements of the specific application: short-insert libraries generate multiple subreads from both strands of the template molecule, which is useful for generating higher accuracy consensus subreads, and long-insert libraries are used to generate the longest possible subreads, which is critical for de novo assembly and detection of structural variants. **b** | SMRT sequencing relies on a sequencing-by-synthesis approach. A DNA polymerase is bound within a zeptolitre-scale observation chamber, called a zero-mode waveguide (ZMW), and uses a strand from the native sequencing library as a template for the read, incorporating fluorescently labelled deoxyribonucleoside triphosphates (dNTPs) as they diffuse into the ZMW. Each incorporated dNTP is briefly immobilized at the polymerase active site, emitting a fluorescent pulse in the corresponding colour channel.

**c** | When observing the fluorescent traces produced by each ZMW, which are highly multiplexed on a chip, the order of pulses provides the read sequence, and pauses between pulses indicate the presence of a covalent modification in the template DNA. **d** | The 1D library preparation from Oxford Nanopore Technologies (ONT) uses a lead adaptor (loaded with a motor protein) and a tethering adaptor, which helps co-locate the molecule near the nanopore, to enable the sequencing of a single DNA strand from the molecule<sup>58</sup>. **e** | ONT sequencing instruments rely on engineered biological nanopores embedded in a lipid membrane to sequence single-stranded DNA (ssDNA). A voltage potential is applied across the membrane, and ssDNA is ratcheted through the nanopore by a motor protein bound to the DNA library molecule. **f** | The ionic current flowing through the nanopore depends on the precise set of nucleotides occupying the constriction point. Methylated nucleotides in the ssDNA introduce distinct current patterns, making it possible to detect the existence of modified bases relative to non-methylated DNA or precomputed models. For clarity, only two changes in current levels are shown in each box.

also detect these modifications in individual reads, albeit at substantially lower sensitivities than when it is applied to consensus current signals from multiple aligned reads. However, the model is constrained by the contents of its training data, which limits its ability to identify novel modification types or novel methylated motifs. Although encouraging, such model-based approaches remain limited in their ability to identify DNA methylation at various sequence contexts. A pre-print article has described an alternative method for nanopore-based methylation detection that uses a statistical comparison of current signals from native and methylation-free WGA DNA<sup>66</sup> and that builds upon the design first proposed for detecting base modifications in SMRT sequencing<sup>27</sup>. This approach is not limited to the detection of DNA methylation at specific sequence motifs and has detected several expected 4mC, 5mC and 6mA motifs in bacterial genomes carrying MTases of known specificity, although the detection accuracy fluctuates with different methylation types and motif specificities<sup>66</sup>. Although encouraging, detection is not yet possible at the level of single molecules, and methods such as this one that do not require a priori knowledge may not be able to distinguish between diverse forms of DNA modification events, especially in eukaryotic genomes<sup>68</sup>.

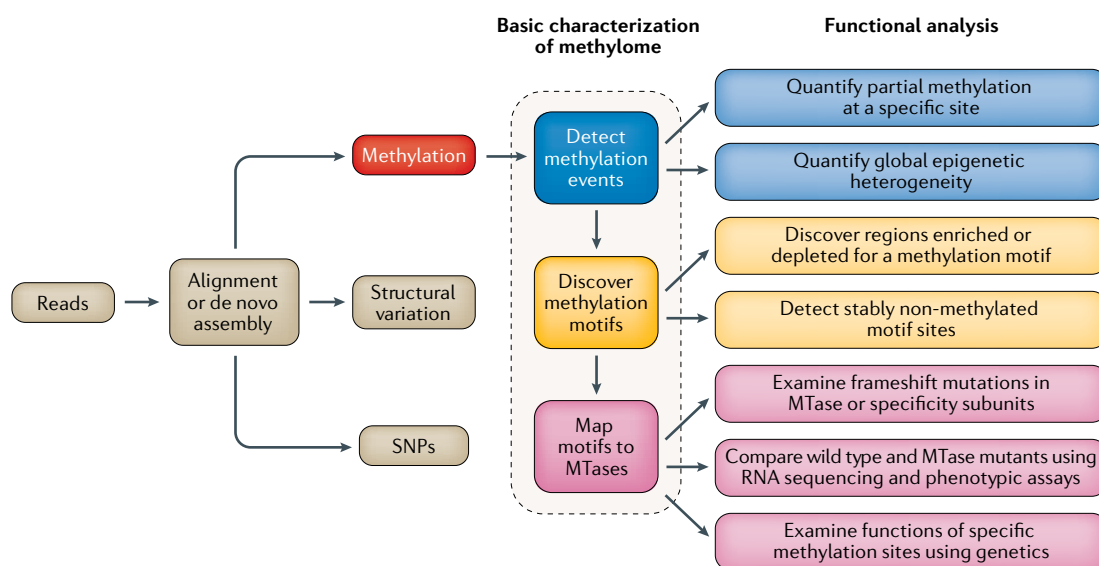
These studies have demonstrated the feasibility of nanopore-based methylation detection; however, some challenges remain. For instance, none of these nanopore-sequencing-based methods has been applied for the biological characterization of an unknown bacterial methylome. Nonetheless, the rapid pace of method development in this field and the ongoing technological advancements in the underlying

sequencing technology make nanopore sequencing a promising field to watch.

### Methylation motifs and methyltransferases

Comprehensive mapping of a bacterial methylome requires more than just the detection of methylated nucleotides; it also requires identification of methylation motifs and the MTases responsible for the observed methylation patterns (FIG. 4).

**Identifying methylation motifs.** DNA methylation events in prokaryotic genomes are highly motif-driven for all three of the primary methylation types. If a methylation motif is targeted by an MTase, typically >95% of motif occurrences are methylated<sup>2,18,20,53</sup>. Historically, methylation motifs for novel type II RM systems have been identified through restriction digest approaches, as in these systems, restriction occurs precisely within the specific sequence motifs. However, the restriction site cannot serve as a proxy for the methylation motif in type I and type III RM systems, in which restriction occurs at a variable distance from the site of methylation<sup>69</sup> (FIG. 2). As a result, there was until recently a notable paucity of known type I and type III RM systems contained in REBASE. However, the introduction of SMRT sequencing and the resulting accumulation of methylome surveys have contributed to a rapidly growing catalogue of known bacterial RM systems<sup>19,29,30,70–72</sup> (FIG. 5). The initial output of a methylation survey is a list of genomic positions that are likely modified. In order to infer the methylation motif from this list, tools such as MEME<sup>73</sup> can be used to build a consensus motif from the sequence context immediately surrounding the modified position.



**Fig. 4 | Steps for comprehensive characterization of a bacterial methylome.** Methylome characterization is increasingly becoming a standard component of bacterial genomic research. The detection of methylated positions can lead to the identification of precise methylated sequence motifs. A methylated motif can then be assigned to the responsible methyltransferase (MTase) based on either querying a database of MTases with known target motifs or through experimental means that involve comparing wild-type strains with strains where the MTase is inactivated. Multiple lines of functional investigation can lead from this basic characterization of the primary features of a bacterial methylome. SNPs, single-nucleotide polymorphisms.

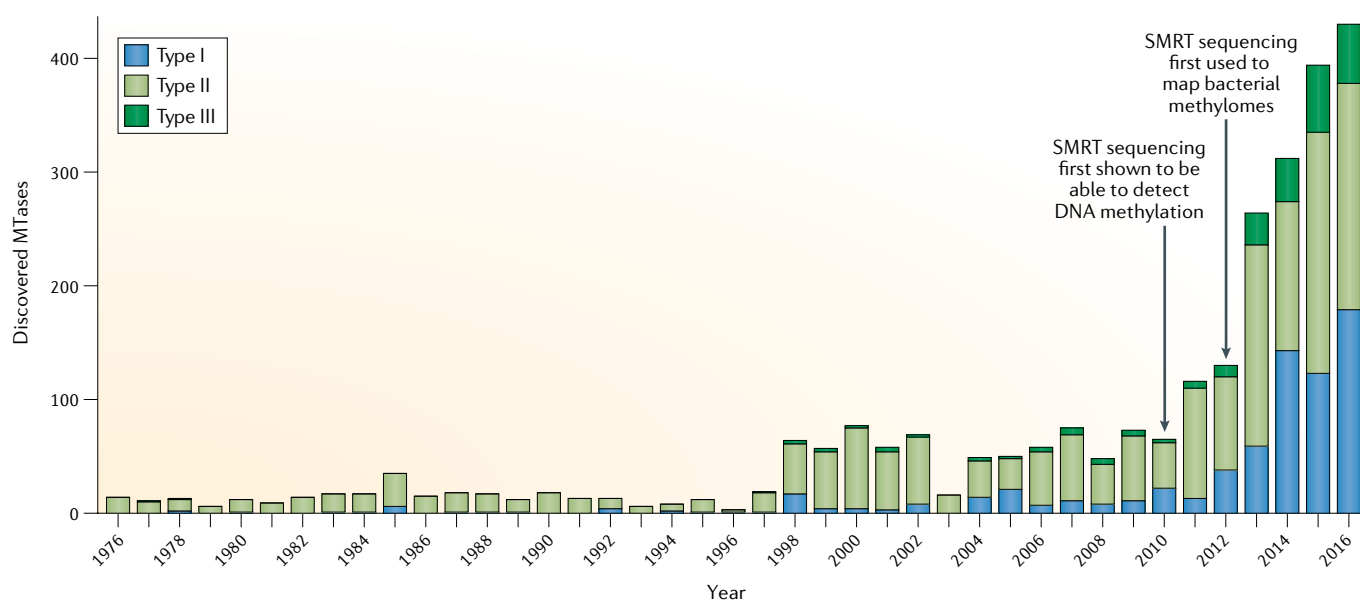


Fig. 5 | **The accelerating pace of methyltransferase discovery.** Historically, type II methyltransferases (MTases) have been the most amenable to discovery, primarily through restriction enzyme digest and fragment analysis. By contrast, restriction enzyme digest is not well suited to de novo discovery of methylated motifs in type I and type III restriction–modification systems because the cut sites of cognate restriction enzymes are typically located at a variable distance from the methylated motif site. The introduction of methylation detection using single-molecule, real-time (SMRT) sequencing in 2010–2012 resulted in a surge of newly discovered MTases belonging to these systems.

#### Identifying methyltransferases responsible for motif methylation.

Gene prediction tools and homology search tools, such as SEQWARE<sup>74</sup>, are often used to identify genes likely to encode components of an RM system, including subunits responsible for restriction, specificity and MTase activity. These components are typically encoded by genes proximal to each other in the genome and can be classified by RM system type (FIG. 2) on the basis of type-specific functional domains. Once classified by type, the characteristic methylation properties of the different RM system types can be leveraged to narrow down the list of putative MTases responsible for an observed methylation motif. For instance, type I MTases target complementary bipartite motifs on both strands, while type III MTases target contiguous, non-palindromic motifs on a single strand. After narrowing the set of candidate MTase genes, the sequences of these candidates can be queried against MTase sequences with known motif specificities in REBASE<sup>19</sup>, and a high-quality sequence match is often sufficient for a confident mapping<sup>20,28</sup>. In the absence of a high-quality MTase match in REBASE, two experimental approaches can be used to identify the MTase responsible for an observed methylated motif. The first relies on heterologous expression of the putative MTase gene in an otherwise non-methylated host, such as *E. coli* ER2796 (REFS<sup>3,28,52,75,76</sup>). Alternatively, the putative MTase can be subjected to an inactivating mutation, where the mutation is either introduced experimentally<sup>29,77</sup> or occurs naturally in a related strain<sup>78,79</sup>. If heterologous expression of the MTase results in methylation of the motif in question or if inactivation of the MTase abolishes methylation at that motif, the causal role of that MTase is confirmed.

#### Insights into restriction–modification systems from methylome analysis.

RM systems often represent an important obstacle to genetic manipulation of an organism by leading to low transformation efficiencies. The design of effective shuttle vectors must, therefore, either include compatible methylation patterns to provide protective methylation or limit the number of motif sites in the vector that are subject to restriction by the host RM system<sup>80,81</sup>. Both of these approaches require a thorough understanding of the host RM repertoire and benefit from a comprehensive catalogue of known RM systems and specificities.

Perhaps the most surprising observation to come from the multitude of prokaryotic methylome studies is the remarkable diversity of MTase genes and target specificities. A recent survey of 230 diverse bacterial and archaeal epigenomes, which was enabled by SMRT sequencing, found DNA methylation in 93% of genomes across a wide diversity of methylated motifs (834 distinct motifs; averaging three motifs per organism)<sup>20</sup>. The primary driver behind this diversity is the spread of MTase-containing mobile genetic elements through horizontal gene transfer (HGT)<sup>20,82,83</sup>. Mutation events can also occur in the target recognition domain of MTase genes and thereby modify the sequence motif targeted for methylation, providing a route to further methylome diversification<sup>30</sup>. As a consequence of such diversification, researchers commonly find substantially divergent methylomes not only among species but also among different strains of the same species<sup>29,72,84–89</sup>.

#### Insights into methylation types

The recent surge of studies devoted to the characterization and functional examination of bacterial methylomes has built upon decades of previous work, most of

which has relied on experimental approaches focused on a handful of loci in a relatively small number of well-characterized genomes<sup>4,10,38,90–92</sup>. The hard-won insights from these foundational studies have long hinted at an unappreciated level of complexity and regulatory potential present in modifications to the four canonical bases. Genome-wide mapping by modern methylation detection technologies is shedding new light on the distribution and roles of the three primary forms of DNA methylation in bacteria.

**N4-methylcytosine.** The extent of 4mC in bacterial genomes is not well understood, and its function largely remains a mystery, although it is known to be involved in multiple RM systems. 4mC occurs less frequently than 6mA in all bacteria but has been observed more often in thermophilic bacteria than in non-thermophilic bacteria, possibly because it is substantially more resistant to heat-induced deamination than 5mC, the other form of cytosine methylation found in bacteria<sup>93,94</sup>. RM-based analysis and modified bisulfite methods have been used to map 4mC in a number of bacterial genomes, including *Caldicellulosiruptor kristjanssonii* and *Enterococcus faecalis*<sup>47,81,95,96</sup>. However, SMRT sequencing is currently the most broadly applied method for 4mC detection. A variety of 4mC motif specificities have been identified in a range of species, including *Bacillus cereus*<sup>28</sup>, *Helicobacter pylori*<sup>29,30,70</sup>, *Campylobacter jejuni*<sup>71</sup> and *S. enterica*<sup>72</sup>. Despite this progress in mapping the distribution of 4mC, its biological functions remain largely unclear. Only a single published study on the gastric carcinogenic bacterium *H. pylori* provides new insight into its potential functions. Deletion of the 4mC MTase M2.HpyAII altered the expression of 102 genes, resulting in decreased adherence to a human gastric adenocarcinoma cell line, reduced potential to induce inflammation and a diminished capacity for natural transformation<sup>97</sup>.

**5-methylcytosine.** Dcm, the orphan cytosine MTase of *E. coli*, has been the subject of study for several decades, and its target specificity of 5'-CCWGG-3' has long been known<sup>98</sup>. The EcoRII RM system is known to protect bacteria against parasitism<sup>99</sup>, and methylation by Dcm has been associated with Tn3 transposition<sup>100</sup>, lambda phage recombination<sup>101</sup> and the expression of ribosomal proteins during stationary phase<sup>102</sup>. However, more general insights into the biological role of 5mC in bacteria have remained somewhat elusive.

Two recent studies have taken advantage of the genome-wide and single-nucleotide resolution of bisulfite sequencing to thoroughly investigate 5mC function in Gammaproteobacteria. In the first study, deletion of Dcm and the resulting suppression of methylation at 5'-CCWGG-3' in *E. coli* resulted in increased expression of the RNA polymerase sigma factor *rpoS* gene and many of its target genes during stationary phase<sup>45</sup>. The second study revealed that methylation of 5'-RCCGGY-3' by the cytosine MTase VchM is required for optimal growth in *Vibrio cholerae* and affects the cell envelope stress response, potentially by downregulating genes required for modifying the lipopolysaccharide

inner core of the cell envelope<sup>46</sup>. While it is tempting to conclude from these studies that 5mC in bacteria is a suppressor of gene expression, more work will be needed to confirm this role — particularly as neither study demonstrated direct regulation of transcription by 5mC methylation.

SMRT sequencing has revealed the 5mC motif specificities of active cytosine MTases in a variety of bacterial species and strains<sup>29,80,84,88,103</sup>. Identification of 5mC methylated positions in isolates of *Neisseria meningitidis* showed them to be mutational hot spots, indicating that 5mC methylation may play a role in genome plasticity and evolution<sup>84</sup>. An improved understanding of 5mC motif specificities has also facilitated the design of plasmids capable of overcoming barriers to transformation in an important strain of *Bifidobacterium animalis*<sup>80</sup>, thereby enabling the molecular mechanisms underlying the observed correlations between bifidobacteria and gut health to be studied<sup>104</sup>.

**N6-methyladenine.** Although traditional RM digestion-based approaches were used in a recent genome-wide mapping study of 6mA<sup>105</sup>, the majority of bacterial studies have adopted SMRT sequencing. The abundance of 6mA MTases in the bacterial world and the robust IPD signature generated by 6mA during SMRT sequencing have led to the discovery of a vast diversity of 6mA MTases and methylated motifs in bacteria, which include many previously unknown orphan MTases and a multitude of previously uncharacterized type I, II and III RM systems<sup>19,20</sup> (FIG. 5). The elucidated 6mA methylated motifs span a wide variety of organisms across multiple phyla, including Bacteroidetes<sup>87</sup>, Firmicutes<sup>81,89,106,107</sup>, Actinobacteria<sup>31,80,85</sup> and Proteobacteria<sup>28–30,70–72,76–78,84,86,88,103,108,109</sup>. Functional knockout studies in many of these organisms highlight the ability of certain 6mA MTases to induce widespread transcriptional changes<sup>3,18,31,106,110,111</sup>, while other work has revealed differentially methylated 6mA positions in response to varied growth stages and environmental conditions<sup>31,77,103</sup>.

Researchers have also taken advantage of SMRT sequencing to explore mechanisms of bacteriophage invasion and host defence. For instance, 936-type bacteriophages, which commonly infect *Lactococcus lactis* starters used in cheese production, have been shown to encode multiple 6mA MTases<sup>75</sup>, which likely provide the bacteriophages with protective methylation that allows them to circumvent host RM systems. Conversely, the bacteriophage exclusion system is a gene cassette that confers bacteriophage resistance in a wide range of host bacteria. Interestingly, although activity of a 6mA MTase in the cassette is required for successful host defence, phage DNA does not appear to be targeted for restriction, which suggests a novel mechanism of methylation-based host defence<sup>112</sup>.

### Insights into epigenetic regulation

In addition to providing a better understanding of the modifications and modifiers involved in DNA methylation, advances in methylation detection methods are also starting to reveal their biological functions and, in



some cases, the mechanisms by which they exert their biological effect.

**Methylation as a cellular regulatory signal.** Several MTases have been shown to be capable of inducing consequential shifts in gene expression<sup>3,45,46,110,113</sup>. For instance, in the local competition model, competitive binding between an MTase and other DNA-binding proteins (such as transcription factors) at specific motif sites affects transcription of a nearby gene, leading to phenotypic variation within the bacterial population<sup>6,114–116</sup> (FIG. 6a). In effect, the MTase methylates specific targets in some fraction of the population, thereby inducing or repressing local transcription in this fraction of the population. Canonical examples of this model in *E. coli* include the transcriptional regulation of the pyelonephritis-associated pili (*pap*) operon and the outer membrane protein-encoding gene *agn43* by Dam methylation at nearby 5'-GATC-3' sites<sup>114,115,117</sup>. In the case of *pap*, which is required for adherence of uropathogenic *E. coli* to the host urinary tract, the local processivity of Dam is hindered by the sequence context of the 5'-GATC-3' in the *pap* promoter, which provides more time for the DNA-binding proteins to access their target sites and thereby skews the competition in their favour<sup>118</sup>. In both examples, methylation provides a means for modulating the antigenic profile of the population, thereby playing a role in immune evasion of host-adapted pathogenic bacteria.

Traditional methylation detection approaches have identified examples of antigenic variation in other Gammaproteobacteria that are generated by competition between Dam and DNA-binding proteins, including in the leucine-responsive regulatory protein (Lrp) and the oxidative stress response protein (OxyR)<sup>2,4,119–122</sup>. However, the prevalence of this type of epigenetic regulation was revealed only when it became possible to systematically map the frequency and distribution of non-methylated sites with SMRT sequencing. Studies have reported several hundred non-methylated motif sites across various bacteria<sup>31,85,103,110,123</sup>; these sites are enriched at gene regulatory regions<sup>2,4,20</sup>, which suggests they are involved in competitive regulation of gene expression. Although detailed mechanisms remain to be identified in most cases, it has recently been shown through SMRT sequencing that variable site-specific Dam methylation at a 5'-GATC-3' motif in the regulatory region of the *opvAB* operon of *S. enterica* is responsible for determining the O-antigen chain length, which is a major determinant of phage resistance in *S. enterica*<sup>124,125</sup>.

DNA methylation also exerts critical regulatory signals during DNA replication. For example, specific motif sites in the replication origins of *E. coli* and *Caulobacter crescentus* (5'-GATC-3' and 5'-GANTC-3', respectively) must be methylated for replication to occur<sup>6</sup>. Furthermore, Dam methylation of a hemimethylated 5'-GATC-3' site in the promoter of the transposase gene of IS10 represses its transcription during replication, presumably to ensure that potential transposition occurs only when a cell contains more than one copy of the chromosome<sup>10</sup>. Clues to the biological function

of an MTase can occasionally be found by identifying genomic regions that are enriched for the methylation target motifs. For instance, enrichment of the Dam 5'-GATC-3' target motif near the origin of replication in *E. coli* and other Gammaproteobacteria has been well documented and is linked to its roles in the initiation of replication<sup>77,92</sup>. SMRT sequencing has revealed further examples of enrichment of 6mA motif sites near origins of replication in *Arthrobacter* and *Nocardia*, which indicates that this phenomenon may not be limited to Gammaproteobacteria<sup>20</sup>. Examples of over-enriched and under-enriched motif sites at other regions of the genome have been identified by SMRT in a wide variety of bacteria<sup>3,20,31,46,103</sup>, which could provide important clues about the biological purpose of the MTases responsible for their methylation.

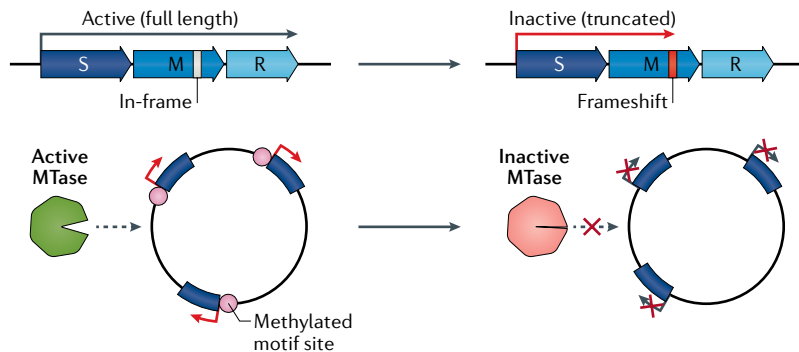
**Phase variation and epigenetic heterogeneity.** Phase variation of bacterial surface proteins, caused by reversible mutations at a hypervariable locus<sup>126–128</sup>, has long been recognized as a mediator of antigenic variation and immune evasion. However, the importance and extent of phase-variable MTases have only more recently become apparent. Hypervariable mutations in the regions regulating and encoding these MTases can result in cell-to-cell differences in their expression (which affects the methylation status of their targets; FIG. 6b) or in their target specificity (which results in methylation of a different set of targets; FIG. 6c). Consequently, heterogeneous methylation patterns can develop within a clonally expanded population, which often has dramatic and genome-wide regulatory consequences<sup>69,129–132</sup>. The set of genes affected by a particular phase-variable MTase is called a phase-variable regulon or a phasevarion<sup>69,130</sup>. Note that this phenomenon is different from the example of phase variation of surface proteins described in the previous section, in which population-level variation in methylation at specific motif sites is caused by competition between MTases and DNA-binding proteins and not by phase-variable MTases.

Phase-variable MTases were first observed almost two decades ago as hypervariable inversion events in the *hsd* genes of *Mycoplasma pulmonis*<sup>133</sup> and *Streptococcus pneumoniae*<sup>134</sup>, which encode the type I RM system of these bacteria. Further examples were subsequently uncovered in *Pasteurella haemolytica*<sup>135</sup>, *Moraxella catarrhalis*<sup>136</sup>, *Haemophilus influenzae*<sup>130,137–139</sup>, *H. pylori*<sup>140–142</sup> and *N. meningitidis*<sup>79,139,143,144</sup>. Their biological importance was quickly appreciated owing to their effects on multiple genes throughout the genome, but in the absence of techniques to determine the underlying motif-specific methylation events, their phase-variable behaviour could be inferred only indirectly, and the precise mechanisms by which they affect gene expression remained unknown. SMRT sequencing has since been used to characterize the target motifs and methylation sites for a number of previously identified phase-variable MTases from a range of bacteria, including ModA and ModD in *N. meningitidis*<sup>109,144</sup>, ModM2 in the human respiratory pathogen *M. catarrhalis*<sup>78</sup> and ModA in *H. influenzae*<sup>76,130,137–139</sup>. It has also provided insights into the mechanisms that give rise to phase-variable

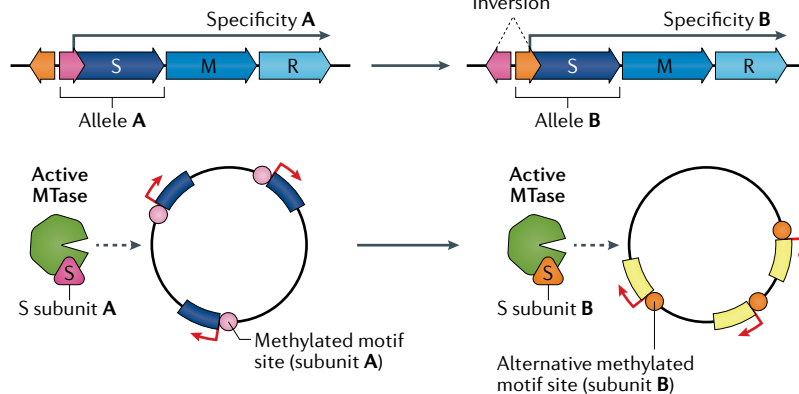
**a Site-specific competitive binding**



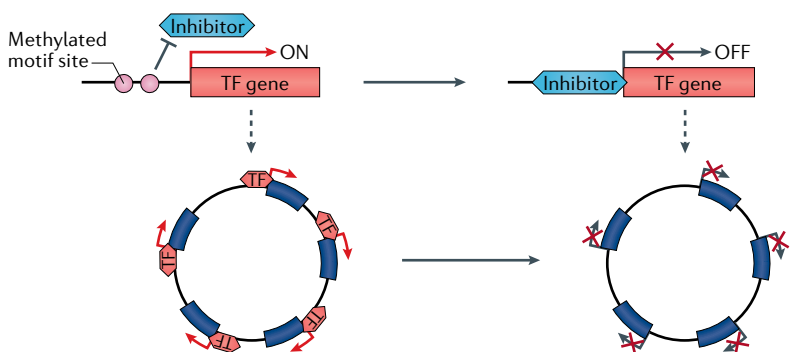
**b MTase ON/OFF switching**



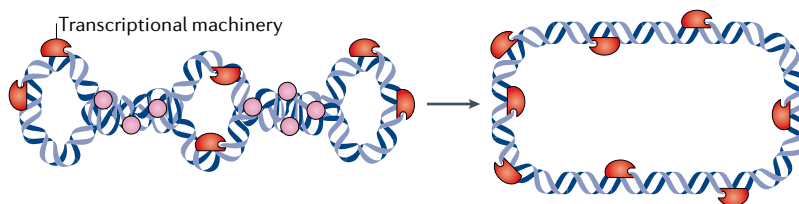
**c MTase specificity switching**



**d TF cascade**



**e Chromosomal conformation**



**Fig. 6 | Epigenetic mechanisms of gene regulation and their consequences.** **a** | The methylation status at motif sites within the upstream regulatory sequence of a gene can affect its expression. The presence of methylated bases in this region can interfere with the binding of regulatory proteins, leading to either upregulation or downregulation of the gene. For instance, methylation can prevent a transcription factor (TF) from binding to its TF-binding site (TFBS), thereby preventing transcription of the downstream gene. **b** | Phase-variable methyltransferases (MTases) are capable of inducing genome-wide changes in methylation status and gene expression. Spontaneous and reversible frameshift mutations, often caused by slipped-strand mispairing in tandem repeat sequences during replication, induce inactivating premature stop codons in the gene encoding the modification (M) subunit. Cells containing the inactivated form of the M subunit lack methylation at the target motif, thereby providing a clonally expanded bacterial population with divergent methylation activity and distinct gene expression regimes<sup>69</sup>. In type I and type III restriction–modification (RM) systems where ON/OFF phase variation is most common, restriction activity requires both restriction (R) subunits and full-length M subunits. Therefore, both the methylation and restriction functions of these RM system are toggled ON and OFF by these frameshift mutations. **c** | Genetic rearrangements, such as inversion events, within the gene encoding the specificity (S) subunit can result in the expression of multiple specificity alleles. When paired with an active M subunit, these diversified S subunits target multiple motifs for methylation<sup>106</sup>. **d** | If a gene affected by methylation status encodes a TF or another protein with promiscuous DNA-binding specificity, the local methylation status can potentially trigger a cascade of downstream changes on gene expression. **e** | DNA methylation is likely to be involved in alternative mechanisms of gene regulation. For example, methylation is known to affect the curvature of DNA molecules, which could potentially control which regions of a chromosome are exposed to the transcriptional machinery of the cell, thereby affecting gene expression.

MTases and how they regulate phasevariations. Studies aiming to characterize how phase-variable MTases in *H. pylori* contribute to its highly complex methylome identified multiple phase-variable MTases generated by slipped-strand mispairing in homopolymer tracts as well as an unusual type I MTase that targets multiple bipartite motifs by interacting with several target recognition domain elements; this process can generate methylome diversification through recombination within the specificity subunit (S subunit)<sup>29,30</sup>. Although it had been previously shown that phase-variable MTases in *H. pylori* are capable of regulating phasevariations<sup>142</sup>, SMRT sequencing demonstrated the importance of the *ModH5* allele of the phase-variable *ModH* MTase in regulating virulence genes in *H. pylori*<sup>145</sup>. A study in *S. pneumoniae* found that a previously observed phase variation of the type I *hsd* system<sup>134</sup> is capable of inducing dramatic changes in the bacterial methylome. This example is one of the most complex phasevariations characterized to date: reconfiguration of five target recognition domains in the S subunit leads to six possible MTase alleles, each with its own target specificity<sup>106</sup>. Taken together, these

findings have deepened our understanding of previously identified phase-variable MTases.

Other studies have taken advantage of the hypothesis-free nature of analysing methylomes with SMRT sequencing to uncover novel phase-variable MTases in other pathogenic bacteria. For example, SMRT sequencing led to the recent discovery of MTase phase variation in the human gastric pathogen *C. jejuni*<sup>108</sup> and the bovine respiratory pathogen *Bibersteinia trehalosi*<sup>88</sup>. The phase variation in *C. jejuni* was shown to affect cell adherence, invasion and biofilm formation, but additional study is required to determine the functional consequences of MTase phase variation in *B. trehalosi*. Use of a software package named SMALR, which was developed to extract single-molecule-level methylation status from SMRT sequencing data, revealed a new type of epigenetic heterogeneity in the marine bacterium *Chromohalobacter salexigens*<sup>18</sup>, in which methylation is dispersed across some, but not all, instances of a target motif. The biological reason for this observed pattern of incomplete methylation is unknown.

There is now a wealth of evidence indicating that MTase phase variation is a crucial survival mechanism for host-adapted bacteria. Variability in methylation patterns has been observed to affect gene expression and phenotypes, but future work will be required to clarify the precise mechanisms through which methylation regulates gene expression.

**Epigenetic regulation of clinically important phenotypes.** Of the many molecular and cellular phenotypes regulated by DNA methylation, clinically important phenotypes are of particular interest. Previous studies using traditional methods hinted at the clinical relevance of bacterial methylation. For instance, methylation of 5'-GATC-3' by Dam in *Salmonella enterica* subsp. *enterica* serovar Typhimurium was shown to be essential for virulence<sup>146,147</sup>. More recent studies have linked additional clinically important phenotypes to bacterial DNA methylation, and many have used SMRT sequencing to associate specific methylation motifs targeted by phase-variable MTases with particular phenotypes.

For instance, the *ModA11* and *ModA12* alleles of the phase-variable ModA MTase in *N. meningitidis* have been linked to sensitivity to several antibiotics that are typically prescribed for meningococcal disease. The phase-variable ModD MTase has also been linked to hypervirulent strains of the same pathogen<sup>79,144</sup>. The phase-variable MTase ModM in *M. catarrhalis* has potential roles in colonization, infection and immune evasion in humans. Specifically, a recent study observed a significant enrichment of the *ModM3* allele over the more common *ModM2* allele in middle ear isolates from individuals with otitis media, which suggests that genes regulated by ModM3 methylation play a part in colonization and infection<sup>78</sup>. Specific *ModA* alleles of *H. influenzae* were selected for in vivo during progression of otitis media in chinchillas, suggesting a role for DNA methylation in *H. influenzae* colonization and infection<sup>76</sup>. Additionally, experiments using locked variants of these phase-variable *ModA* alleles demonstrated regulation of a variety of clinically important pathways, such as

immune evasion, biofilm formation, antibiotic susceptibility, virulence and niche adaptation<sup>76,148</sup>. These results corroborate orthogonal studies that implicate ModA phase variation as an important regulator of virulence and immune evasion<sup>149,150</sup>. In *S. pneumoniae*, the MTase of the SpnD39III RM system possesses six specificity alleles that are generated through genetic rearrangement and that target different motifs for methylation. These alleles have different virulence phenotypes and are selected at various stages of colonization and infection<sup>106</sup>.

Collectively, these studies imply that many bacterial pathogens exploit epigenetic switches as a flexible mechanism to regulate gene expression during host colonization and infection. Some of these mechanisms may serve as targets of potential therapeutic intervention strategies.

**Towards deeper mechanistic insights.** The first step in studying the functional impacts of bacterial DNA methylation is to compare global gene expression between wild-type strains and MTase mutant strains. A number of studies that used RNA sequencing for such comparisons have shown that perturbation of a single DNA MTase often results in differential expression of tens or hundreds of genes, reaching as many as a thousand in some cases<sup>3,18,31,106,110,111</sup>. These data highlight that the effects of DNA methylation in the regulation of bacterial gene expression have been underestimated but also reveal some unexpected findings. In some cases, the regulatory effects of MTases can be conclusively traced to methylation at the promoters of target genes. For instance, the ModH5 MTase in *H. pylori* has been shown to regulate the activity of the gene flagellin A (*flaA*) via methylation in the *flaA* promoter<sup>145</sup>. However, generally, only a small proportion (<10%) of the differentially expressed genes have methylated sites in their promoter regions<sup>3,45,46,110</sup>, which implies that the local competition model, in which a DNA MTase and other DNA-binding proteins compete for binding at the promoter of a gene, does not apply to most differentially expressed genes (FIG. 6a). Another possibility is that the methylation status at individual motif sites might regulate the expression of a transcription factor, causing a broad downstream shift in the expression of its target genes (FIG. 6d). In order to determine which mechanisms are at work, specific methylation sites must be individually mutated using genetic tools such as site-directed mutagenesis<sup>114–116</sup>. Multiple studies have observed a positive correlation between the number of methylation sites in a gene and the fold change of expression between wild type and MTase mutants<sup>3,46</sup>, suggesting that epigenetic regulation of expression may in fact be driven by multiple methylation sites in both the promoter region and the gene body. Another intriguing hypothesis relates to the effect of DNA methylation on the chromosome topology<sup>151–153</sup>, whereby methylation induces structural changes that alter the repertoire of genes exposed to the cellular transcriptional machinery (FIG. 6e).

#### Comparisons with eukaryotic methylomes

Analyses of DNA methylation in eukaryotic genomes have focused on 5mC. However, even with the advent of second-generation and third-generation sequencing

technologies, functional studies of 5mC in the bacterial kingdom have been rare because it is less prevalent than 6mA. Thus, comparisons between bacterial and eukaryotic methylomes have not been feasible until the recent discovery of 6mA in a number of eukaryotes<sup>154</sup>, including algae<sup>155</sup>, fungi<sup>156</sup>, worms<sup>157</sup>, insects<sup>158</sup> and mammals<sup>159,160</sup>. These studies have revealed diverse functions for eukaryotic 6mA, including the regulation of gene expression<sup>157,158,160</sup>, regulation of transposon mobility<sup>158,160</sup> and crosstalk with histone variants and histone modifications<sup>157,160</sup>.

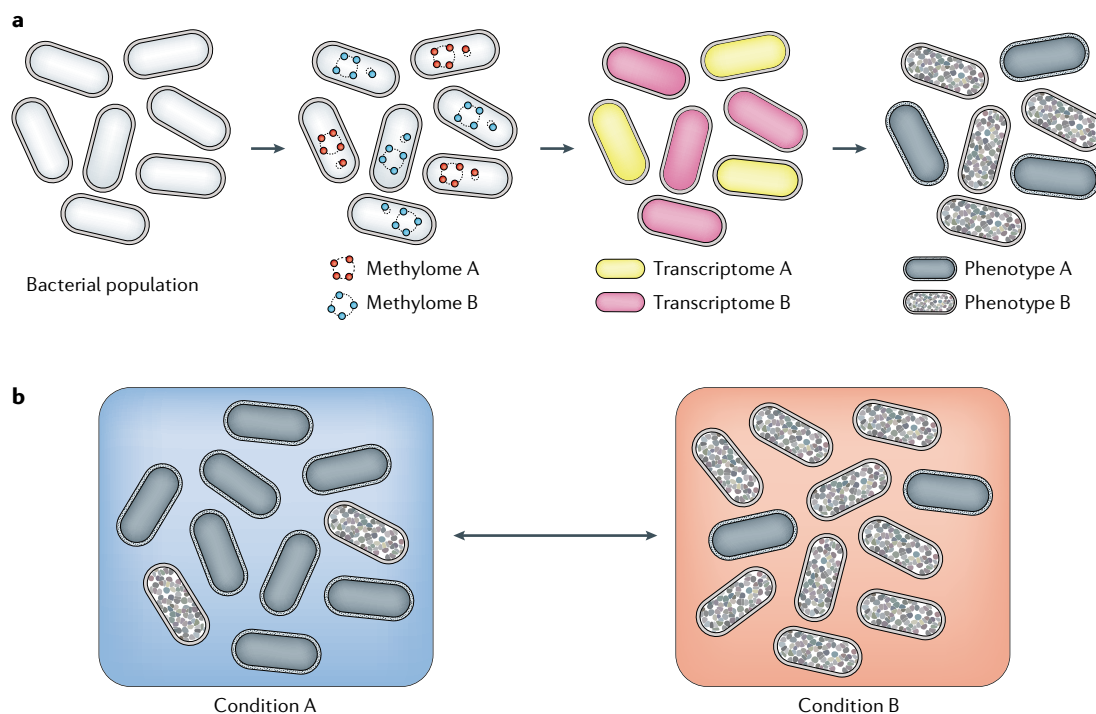
The genomic distribution of 6mA differs between prokaryotes and eukaryotes. The frequency of 6mA (as a fraction of the total number of adenine residues in the genome) is orders of magnitude lower in most eukaryotes than in prokaryotes<sup>68,154</sup>. Furthermore, eukaryotic 6mA events are much less motif-driven than those in prokaryotes. For example, very few occurrences (often < 3%) of 6mA motifs identified in the genomes of *Chlamydomonas reinhardtii*, *Caenorhabditis elegans*, *Plasmodium falciparum* and mouse embryonic stem cells (mESCs) are methylated<sup>68,155,157,160,161</sup>. One likely explanation for these observations is that modified 6mA sites are not targeted by cognate restriction enzymes in eukaryotes and, therefore, do not need to be located at specific sequence motifs. Another possible reason is that MTases have limited access to eukaryotic DNA because it is packaged in nucleosomes, and thus only exposed motifs will be methylated.

Despite these important differences, some commonalities do exist. For example, 6mA events are known to

inhibit the transcription of a form of transposon called insertion elements in bacteria<sup>10,162</sup>, which is analogous to the observed enrichment of 6mA events at transposons in both *C. elegans* and mESCs<sup>157,160</sup>. More fundamentally, the intrinsic properties of 6mA and its effect on DNA conformation are expected to be consistent between bacteria and eukaryotes<sup>151</sup>, although different organisms may exploit these properties in different molecular and cellular contexts. Complete high-resolution maps will be the foundation for future comparisons of bacterial and eukaryotic 6mA methylomes. Although SMRT sequencing and Oxford Nanopore sequencing hold great promise for mapping DNA methylation in bacteria, their successful application to eukaryotic genomes faces critical challenges stemming from the scarcity of modified sites and the lack of clear target motifs. As recent work has suggested, 6mA detection in eukaryotes requires cross-validation by integrating complementary sequencing technologies with molecular technologies based on restriction enzymes and 6mA antibodies<sup>68</sup>.

### Future perspectives

**Integration with orthogonal assays for mechanistic insights.** Technological breakthroughs have made it easier than ever to map bacterial methylomes. However, comprehensive studies will be necessary to fully characterize the precise mechanisms by which DNA methylation modulates gene expression and alters bacterial phenotypes. Such studies would benefit from a richer collection of functional genomics data (such as transcription factor binding assays) from many bacterial



**Fig. 7 | Phenotypic consequences of epigenetic heterogeneity. a** | The presence of phase-variable methyltransferases (MTases) can introduce heterogeneous methylation patterns in a clonally expanded bacterial population, leading to subpopulations with distinct gene expression regimes and phenotypes. **b** | Phenotypically distinct subpopulations can emerge within a colony as a result not only of genetic variation but also of epigenetic variation, that is, variation in DNA methylation status. These subpopulations serve as units of adaptive selection and provide a means of population-level flexibility in response to rapidly changing environments.



## Chromatin conformation capture

A technique used to assess the spatial organization of chromosomes within a cell. Briefly, DNA is first chemically crosslinked and fragmented. The crosslinked fragments are then ligated. When sequenced, the ligated fragments produce concatemers that help reveal which regions of sequence co-locate within the cell.

species, across different genetic backgrounds (wild type and MTase mutants) and in conditions of growth or stress. These experiments must be followed by genetic experiments that mutate and characterize specific methylation sites. In addition, future studies could test the hypothesis that the thermodynamic effect of DNA methylation induces conformational changes to a bacterial chromosome, rendering previously inaccessible genes accessible to the transcriptional machinery<sup>151,163</sup> (FIG. 6e). Chromatin conformation capture analyses, such as Hi-C, can be used to elucidate the effects of bacterial DNA methylation on DNA conformation and, consequently, on gene transcription<sup>1,54</sup>.

**Phasevarions in vaccine development.** The ability of phase-variable MTases to activate antigenic diversity in host-adapted pathogens (FIG. 7) makes them very relevant for vaccine development. Highly diverse and variable antigens do not make good vaccine candidates and are typically avoided. However, genes for outer membrane proteins or other antigens that lack simple tandem repeats (which are common indicators of phase variation) might still be subject to variable expression if they are part of a phasevarion<sup>164</sup>. Indeed, it has been shown that multiple vaccine candidates are likely subject to this epigenetic means of antigenic variation. Thus, identifying phase-variable MTases and their phasevarions in host-adapted pathogens<sup>76</sup> is likely to facilitate more effective vaccine development.

**Methylation detection using nanopore sequencing.** SMRT sequencing has been instrumental in enabling the study of bacterial methylomes, but other sequencing technologies, such as those commercialized by Oxford Nanopore Technologies, have the potential to make important contributions to the field of bacterial epigenetics in the near future. Assuming continued maturation of the technology and improvements in the modification detection algorithms, the very long read lengths offered by nanopore sequencing devices may

provide single-molecule phased detection of bacterial DNA methylation in samples from a variety of environments. This ability will be helpful for the epigenetic study of heterogeneous bacterial samples, including metagenomic populations, where the study of methylation has so far been limited<sup>18,87</sup>. The recent use of methylation signatures as discriminative features for metagenomic binning suggests that the applications of methylation detection in long reads extend beyond identifying methylated motifs in bacteria<sup>53</sup>.

These advances come at a time when the presence and importance of DNA methylation types that have traditionally been recognized only in prokaryotes, such as 6mA, are being investigated in eukaryotes<sup>156,157,160</sup>. As these epigenetic marks become better understood, it will be interesting to see whether eukaryotic modifications share any functional traits with those found in their prokaryotic ancestors.

## Conclusions

The study of bacterial methylomes has been revolutionized by the introduction of technologies capable of detecting 4mC, 5mC and 6mA at a genome-wide scale and single-nucleotide resolution. Application of these new technologies has led to a greater appreciation for the sheer quantity and diversity of methylation systems and their target specificities in bacteria. Deposition of newly discovered MTase genes and their target motifs to community databases such as REBASE<sup>19</sup> has created a powerful resource for researchers, providing a catalogue of the RM systems that can act as barriers to efficient transformation. Technological advances have also highlighted hypervariable MTases and their consequences on genome-wide methylation, gene expression and phenotypic plasticity. Given the modern sequencing-based tools at their disposal, researchers are better equipped than ever before to probe the previously hidden epigenetic mechanisms of the bacterial realm.

Published online 13 December 2018

- Boyer, H. Genetic control of restriction and modification in *Escherichia coli*. *J. Bacteriol.* **88**, 1652–1660 (1964).
- Casadesús, J. & Low, D. Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* **70**, 830–856 (2006).  
**This review discusses how bacterial DNA methylation acts as a regulatory signal in various bacteria.**
- Fang, G. et al. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* **30**, 1232–1239 (2012).  
**This paper describes the first application of SMRT sequencing to detect bacterial 6mA events at single-base resolution and genome-wide scale.**
- Wion, D. & Casadesús, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* **4**, 183–192 (2006).  
**This review discusses the use of 6mA as a regulatory signal in various bacteria.**
- Løbner-Olesen, A., Marinus, M. G. & Hansen, F. G. Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proc. Natl Acad. Sci. USA* **100**, 4672–4677 (2003).
- Low, D. & Casadesús, J. Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol.* **11**, 106–112 (2008).
- Boye, E., Løbner-Olesen, A. & Skarstad, K. Limiting DNA replication to once and only once. *EMBO Rep.* **1**, 479–483 (2000).
- Boye, E., Stokke, T., Kleckner, N. & Skarstad, K. Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc. Natl Acad. Sci. USA* **93**, 12206–12211 (1996).
- Hsieh, P. Molecular mechanisms of DNA mismatch repair. *Mutat. Res.* **486**, 71–87 (2001).
- Roberts, D., Hoopes, B. C., McClure, W. R. & Kleckner, N. IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**, 117–130 (1985).
- Hernday, A., Krabbe, M., Braaten, B. & Low, D. Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl Acad. Sci. USA* **99**, (Suppl. 4), 16470–16476 (2002).
- Waldron, D. E., Owen, P. & Dorman, C. J. Competitive interaction of the OxyR DNA-binding protein and the Dam methylase at the antigen 43 gene regulatory region in *Escherichia coli*. *Mol. Microbiol.* **44**, 509–520 (2002).
- Bickle, T. A. & Kruger, D. H. Biology of DNA restriction. *Microbiol. Rev.* **57**, 434–450 (1993).
- Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A. & Wilson, G. G. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* **42**, 20–44 (2014).
- Pingoud, A., Wilson, G. G. & Wendt, W. Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* **42**, 7489–7527 (2014).
- Rao, D. N., Dryden, D. T. F. & Bheemanaik, S. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* **42**, 45–55 (2014).
- Furuta, Y. & Kobayashi, I. Mobility of DNA sequence recognition domains in DNA methyltransferases suggests epigenetics-driven adaptive evolution. *Mob. Genet. Elements* **2**, 292–296 (2012).
- Beaulaurier, J. et al. Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat. Commun.* **6**, 7438 (2015).
- Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298–D299 (2015).  
**This paper describes the REBASE database, which has become a central repository for bacterial methylome information.**
- Blow, M. J. et al. The epigenomic landscape of prokaryotes. *PLOS Genet.* **12**, e1005854 (2016).  
**This study describes a comprehensive survey of the methylomes of 230 bacteria, describing the diversity of MTases and specificities.**
- Davis, B. M., Chao, M. C. & Waldor, M. K. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Curr. Opin. Microbiol.* **16**, 192–198 (2013).
- Plongthongkum, N., Diep, D. H. & Zhang, K. Advances in the profiling of DNA modifications: cytosine



- methylation and beyond. *Nat. Rev. Genet.* **15**, 647–661 (2014).
23. Bock, C. Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* **13**, 705–719 (2012).
24. Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **11**, 191–203 (2010).
25. Hirst, M. & Marra, M. A. Next generation sequencing based approaches to epigenomics. *Brief. Funct. Genom.* **9**, 455–465 (2010).
26. Eid, J. et al. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
- This paper is a good introduction to the concepts and technology underpinning SMRT sequencing.**
27. Flusberg, B. A. et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
- This paper provides an early description of 5mC, 5hmC and 6mA detection using SMRT sequencing.**
28. Murray, I. A. et al. The methylomes of six bacteria. *Nucleic Acids Res.* **40**, 11450–11462 (2012).
29. Krebs, J. et al. The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* **42**, 2415–2432 (2014).
- This paper describes how the application of SMRT sequencing to multiple *H. pylori* strains revealed unexpectedly complex methylomes and many novel methylation motifs.**
30. Furuta, Y. et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLOS Genet.* **10**, e1004272 (2014).
31. Lluch-Senar, M. et al. Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution. *PLOS Genet.* **9**, e1003191 (2013).
32. Korlach, J. & Turner, S. W. Going beyond five bases in DNA sequencing. *Curr. Opin. Struct. Biol.* **22**, 251–261 (2012).
33. Sánchez-Romero, M. A., Cota, I. & Casadesús, J. DNA methylation in bacteria: from the methyl group to the methylome. *Curr. Opin. Microbiol.* **25**, 9–16 (2015).
34. Casadesús, J. in *DNA Methyltransferases - Role and Function* (eds Jeltsch, A. & Jurkowska, R. Z.) 35–61 (Springer International Publishing, 2016).
35. Razin, A. & Riggs, A. D. DNA methylation and gene function. *Science* **210**, 604–610 (1980).
36. Robertson, K. D. DNA methylation and human disease. *Nat. Rev. Genet.* **6**, 597–610 (2005).
37. Ito, S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
38. Zweiger, G., Marczyński, G. & Shapiro, L. A. Caulobacter DNA methyltransferase that functions only in the predivisional cell. *J. Mol. Biol.* **235**, 472–485 (1994).
39. Nelson, M., Raschke, E. & McClelland, M. Effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* **21**, 3139–3154 (1993).
40. Rao, B. S. & Buckler-White, A. Direct visualization of site-specific and strand-specific DNA methylation patterns in automated DNA sequencing data. *Nucleic Acids Res.* **26**, 2505–2507 (1998).
41. Bart, A., van Passel, M. W., van Amsterdam, K. & van der Ende, A. Direct detection of methylation in genomic DNA. *Nucleic Acids Res.* **33**, e124 (2005).
42. Broadbent, S. E., Balbontin, R., Casadesús, J., Marinus, M. G. & van der Woude, M. YhdJ, a nonessential CcrM-like DNA methyltransferase of *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* **189**, 4325–4327 (2007).
43. Shell, S. S. et al. DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. *PLOS Pathog.* **9**, e1003419 (2013).
44. Bart, A., Pannekoek, Y., Dankert, J. & van der Ende, A. NmeSI restriction-modification system identified by representational difference analysis of a hypervirulent *Neisseria meningitidis* strain. *Infect. Immun.* **69**, 1816–1820 (2001).
45. Kahramanoglou, C. et al. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat. Commun.* **3**, 886 (2012).
- This paper describes one of the first applications of bisulfite sequencing to characterize 5mC in bacteria.**
46. Chao, M. C. et al. A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen. *PLOS Genet.* **11**, 1–24 (2015).
47. Yu, M. et al. Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. *Nucleic Acids Res.* **43**, 1–10 (2015).
48. Schadt, E. E. et al. Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Res.* **23**, 129–141 (2013).
49. Levene, M. J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).
50. Clark, T. A., Spittle, K. E., Turner, S. W. & Korlach, J. Direct detection and sequencing of damaged DNA bases. *Genome Integr.* **2**, 10 (2011).
51. Clark, T. A. et al. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.* **11**, 4 (2013).
52. Clark, T. A. et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* **40**, e29 (2012).
53. Beaulaurier, J. et al. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nat. Biotechnol.* **36**, 61–69 (2018).
54. Clarke, J. et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
- This paper provides an early description of a protein nanopore with covalently attached adapter continuously differentiating between the four canonical bases and 5mC.**
55. Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
56. Laszlo, A. H. et al. Decoding long nanopore sequencing reads of natural DNA. *Nat. Biotechnol.* **32**, 829–834 (2014).
57. Manrao, E. A., Derrington, I. M., Pavlenok, M., Niederweis, M. & Gundlach, J. H. Nucleotide discrimination with DNA immobilized in the MSPA nanopore. *PLOS ONE* **6**, e25723 (2011).
58. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
59. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
60. Ip, C. L. C. et al. MinION Analysis and Reference Consortium: phase 1 data release and analysis. *F1000Research* **4**, 1075 (2015).
61. de Lannoy, C., de Ridder, D. & Risse, J. A sequencer coming of age: de novo genome assembly using MinION reads. *F1000Research* **6**, 1083 (2017).
62. Laszlo, A. H. et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl Acad. Sci. USA* **110**, 18904–18909 (2013).
- This paper uses a phi29 polymerase to ratchet ssDNA through a protein nanopore and identifies the presence of 5mC and 5hmC in single DNA molecules.**
63. Wescoe, Z. L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J. Am. Chem. Soc.* **136**, 16582–16587 (2014).
64. Rand, A. C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).
- This paper describes the SignalAlign tool, which can detect multiple cytosine modifications and 6mA from nanopore sequencing data.**
65. McIntyre, A. B. R. et al. Nanopore detection of bacterial DNA base modifications. Preprint at *bioRxiv* <https://doi.org/10.1101/127100> (2017).
66. Stoiber, M. H. et al. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. Preprint at *bioRxiv* <https://doi.org/10.1101/094672> (2016).
67. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
68. Zhu, S. et al. Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. *Genome Res.* **28**, 1067–1078 (2018).
- This article emphasizes the challenges and caveats in the use of SMRT sequencing for 6mA detection in eukaryotes. Similar challenges apply to nanopore sequencing.**
69. Srikantha, Y. N., Fox, K. L. & Jennings, M. P. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat. Rev. Microbiol.* **8**, 196–206 (2010).
70. Lee, W. C. et al. The complete methylome of *Helicobacter pylori* UM032. *BMC Genomics* **16**, 424 (2015).
71. O’Loughlin, J. L. et al. Analysis of the *Campylobacter jejuni* genome by SMRT DNA sequencing identifies restriction-modification motifs. *PLOS ONE* **10**, e0118533 (2015).
72. Pirone-Davies, C. et al. Genome-wide methylation patterns in *Salmonella enterica* subsp. *enterica* Serovars. *PLOS ONE* **10**, e0123639 (2015).
73. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, (Suppl. 2), W202–W208 (2009).
74. O’Connor, B. D., Merriman, B. & Nelson, S. F. SeqWare query engine: storing and searching sequence data in the cloud. *BMC Bioinformatics* **11**, S2 (2010).
75. Murphy, J. et al. Methyltransferases acquired by lactococcal 936-type phage provide protection against restriction endonuclease activity. *BMC Genomics* **15**, 831 (2014).
76. Atack, J. M. et al. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nat. Commun.* **6**, 7828 (2015).
- In this paper, SMRT sequencing is used to characterize ON/OFF switching of a type III RM system and its effect on immunoevasion and niche adaptation in an animal model.**
77. Bendall, M. L. et al. Exploring the roles of DNA methylation in the metal-reducing bacterium *Shewanella oneidensis* MR-1. *J. Bacteriol.* **195**, 4966–4974 (2013).
78. Blakeway, L. V. et al. ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* phasevarions in otitis media. *FASEB J.* **28**, 5197–5207 (2014).
79. Seib, K. L. et al. A novel epigenetic regulator associated with the hypervirulent *Neisseria meningitidis* clonal complex 41/44. *FASEB J.* **25**, 3622–3633 (2011).
80. O’Connell Motherway, M. et al. Identification of restriction-modification systems of *Biifidobacterium animalis* subsp. *lactis* CNCM I-2494 by SMRT sequencing and associated methylome analysis. *PLOS ONE* **9**, e94875 (2014).
81. Huo, W., Adams, H. M., Zhang, M. Q. & Palmer, K. L. Genome modification in *Enterococcus faecalis* OG1RF assessed by bisulfite sequencing and single-molecule real-time sequencing. *J. Bacteriol.* **197**, 1939–1951 (2015).
82. Kobayashi, I., Nobusato, A., Kobayashi-Takahashi, N. & Uchiyama, I. Shaping the genome — restriction-modification systems as mobile genetic elements. *Curr. Opin. Genet. Dev.* **9**, 649–656 (1999).
83. Conlan, S. et al. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci. Transl. Med.* **6**, 254ra126 (2014).
84. Sater, M. R. A. et al. DNA Methylation assessed by SMRT sequencing is linked to mutations in *Neisseria meningitidis* isolates. *PLOS ONE* **10**, e0144612 (2015).
85. Zhu, L. et al. Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. *Nucleic Acids Res.* **44**, 730–743 (2016).
86. Mou, K. T. et al. A comparative analysis of methylome profiles of *Campylobacter jejuni* sheep abortion isolate and gastroenteric strains using PacBio data. *Front. Microbiol.* **5**, 782 (2014).
87. Leonard, M. T. et al. The methylome of the gut microbiome: disparate Dam methylation patterns in intestinal *Bacteroides dorei*. *Front. Microbiol.* **5**, 361 (2014).
88. Anton, B. P., Harhay, G. P., Smith, T. P. L., Blom, J. & Roberts, R. J. Comparative methylome analysis of the occasional ruminant respiratory pathogen *Bibersteinia trehalosi*. *PLOS ONE* **11**, e0161499 (2016).
89. Chen, P. et al. Comparative genomics reveals the diversity of restriction-modification systems and DNA methylation sites in *Listeria monocytogenes*. *Appl. Environ. Microbiol.* **83**, e02091–16 (2017).
90. Blyn, L. B., Braaten, B. A. & Low, D. A. Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. *EMBO J.* **9**, 4045–4054 (1990).
91. Boyer, H. W. DNA restriction and modification mechanisms in bacteria. *Annu. Rev. Microbiol.* **25**, 153–176 (1971).

92. Løbner-Olesen, A., Skovgaard, O. & Marinus, M. G. Dam methylation: coordinating cellular processes. *Curr. Opin. Microbiol.* **8**, 154–160 (2005).
93. Ehrlich, M. et al. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N5-methyladenine. *Nucleic Acids Res.* **13**, 1399–1412 (1985).
94. Ehrlich, M., Wilson, G. G., Kuo, K. C. & Gehrke, C. W. N4-methylcytosine as a minor base in bacterial DNA. *J. Bacteriol.* **169**, 939–943 (1987).
95. Chung, D., Farkas, J., Huddleston, J. R., Olivar, E. & Westpheling, J. Methylation by a unique Ssa-class N4-Cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. *PLOS ONE* **7**, e43844 (2012).
96. Vilkaitis, G. & Klimasauskas, S. Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. *Anal. Biochem.* **271**, 116–119 (1999).
97. Kumar, S. et al. N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. *Nucleic Acids Res.* **46**, 3429–3445 (2018).
98. Boyer, H. W., Chow, L. T., Dugaiczky, A., Hedgpeth, J. & Goodman, H. M. DNA substrate site for the EcoRII restriction endonuclease and modification methylase. *Nat. New Biol.* **244**, 40–43 (1973).
99. Takahashi, N., Naito, Y., Handa, N. & Kobayashi, I. A. DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. *J. Bacteriol.* **184**, 6100–6108 (2002).
100. Yang, M. K., Ser, S. C. & Lee, C. H. Involvement of *E. coli* dcm methylase in Tn3 transposition. *Proc. Natl. Sci. Counc. Repub. China. B* **13**, 276–283 (1989).
101. Korba, B. E. & Hays, J. B. Partially deficient methylation of cytosine in DNA at CCATGG sites stimulates genetic recombination of bacteriophage lambda. *Cell* **28**, 531–541 (1982).
102. Millett, K. T. et al. Conservation of Dcm-mediated cytosine DNA methylation in *Escherichia coli*. *FEMS Microbiol. Lett.* **328**, 78–85 (2012).
103. Kozdon, J. B. et al. Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. *Proc. Natl. Acad. Sci. USA* **110**, E4658–E4667 (2013).
104. O'Callaghan, A. & van Sinderen, D. Bifidobacteria and their role as members of the human gut microbiota. *Front. Microbiol.* **7**, 925 (2016).
105. Dalia, A. B., Lazinski, D. W. & Camilli, A. Characterization of undermethylated sites in *Vibrio cholerae*. *J. Bacteriol.* **195**, 2389–2399 (2013).
106. Manso, A. S. et al. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat. Commun.* **5**, 5055 (2014). **In this article, SMRT sequencing is used to characterize a specificity switching RM system involved in colonization and virulence of *S. pneumoniae*.**
107. Li, J. et al. Epigenetic switch driven by DNA inversions dictates phase variation in *Streptococcus pneumoniae*. *PLOS Pathog.* **12**, e1005762 (2016).
108. Anjum, A. et al. 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**, 4581–4594 (2016).
109. Seib, K. L. et al. Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N6-adenine DNA methyltransferases of *Neisseria meningitidis*. *Nucleic Acids Res.* **43**, 4150–4162 (2015).
110. Gonzalez, D., Kozdon, J. B., McAdams, H. H., Shapiro, L. & Collier, J. The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach. *Nucleic Acids Res.* **42**, 3720–3735 (2014).
111. Zhou, B. et al. The global regulatory architecture of transcription during the *caulobacter* cell cycle. *PLOS Genet.* **11**, e1004831 (2015).
112. Goldfarb, T. et al. BREX is a novel phase resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2014).
113. Balbontin, R. et al. DNA adenine methylation regulates virulence gene expression in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **188**, 8160–8168 (2006).
114. der Woude, M. W. Van, Braaten, B. & Low, D. Epigenetic phase variation of the pap operon in *Escherichia coli*. *Trends Microbiol.* **4**, 5–9 (1996).
115. Wallecha, A., Munster, V., Correnti, J., Chan, T. & van der Woude, M. Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. *J. Bacteriol.* **184**, 3338–3347 (2002).
116. Lim, H. N. & Van Oudenaarden, A. A multistep epigenetic switch enables the stable inheritance of DNA methylation states. *Nat. Genet.* **39**, 269–275 (2007).
117. Casadesús, J. & Low, D. A. Programmed heterogeneity: epigenetic mechanisms in bacteria. *J. Biol. Chem.* **288**, 13929–13935 (2013).
118. Peterson, S. N. & Reich, N. O. GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence pap expression. *J. Mol. Biol.* **355**, 459–472 (2006).
119. Davies, M. R., Broadbent, S. E., Harris, S. R., Thomson, N. R. & van der Woude, M. W. Horizontally acquired glycosyltransferase operons drive salmonellae lipopolysaccharide diversity. *PLOS Genet.* **9**, e1003568 (2013).
120. Broadbent, S. E., Davies, M. R. & van der Woude, M. W. Phase variation controls expression of *Salmonella* lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. *Mol. Microbiol.* **77**, 337–353 (2010).
121. Cota, I., Blanc-Potard, A. B. & Casadesús, J. STM2209-STM2208 (opvAB): a phase variation locus of *Salmonella enterica* involved in control of O-antigen chain length. *PLOS ONE* **7**, e36863 (2012).
122. Camacho, E. M. & Casadesús, J. Regulation of traJ transcription in the *Salmonella* virulence plasmid by strand-specific DNA adenine hemimethylation. *Mol. Microbiol.* **57**, 1700–1718 (2005).
123. Cohen, N. R. et al. A role for the bacterial GATC methylome in antibiotic stress survival. *Nat. Genet.* **48**, 581–586 (2016).
124. Cota, I. et al. OxyR-dependent formation of DNA methylation patterns in OpvAB<sup>OFF</sup> and OpvAB<sup>ON</sup> cell lineages of *Salmonella enterica*. *Nucleic Acids Res.* **44**, 3595–3609 (2016).
125. Cota, I. et al. Epigenetic control of *Salmonella enterica* O-antigen chain length: a tradeoff between virulence and bacteriophage resistance. *PLOS Genet.* **11**, e1005667 (2015).
126. Jennings, M. P., Hood, D. W., Peak, I. R., Virji, M. & Moxon, E. R. Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. *Mol. Microbiol.* **18**, 729–740 (1995).
127. van der Ende, A. et al. Variable expression of class I outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J. Bacteriol.* **177**, 2475–2480 (1995).
128. Cerdeño-Tarraga, A. & Patrick, S. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463–1466 (2005).
129. Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches - the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33**, 919–932 (1999).
130. Srihanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M. & Jennings, M. P. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc. Natl. Acad. Sci. USA* **102**, 5547–5551 (2005). **This paper introduces the concept of a phase-variable regulon (phasevarion).**
131. Atack, J. M., Tan, A., Bakaletz, L. O., Jennings, M. P. & Seib, K. L. Phasevarions of bacterial pathogens: methylomics sheds new light on old enemies. *Trends Microbiol.* **26**, 715–726 (2018).
132. Atack, J. M., Yang, Y., Seib, K. L., Zhou, Y. & Jennings, M. P. A survey of type III restriction-modification systems reveals numerous, novel epigenetic regulators controlling phase-variable regulons; phasevarions. *Nucleic Acids Res.* **46**, 3532–3542 (2018).
133. Dybvig, K., Sitaraman, R. & French, C. T. A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc. Natl. Acad. Sci. USA* **95**, 13923–13928 (1998).
134. Tettelin, H. et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506 (2001).
135. Ryan, K. A. & Lo, R. Y. Characterization of a CACAG pentanucleotide repeat in *Pasteurella haemolytica* and its possible role in modulation of a novel type III restriction-modification system. *Nucleic Acids Res.* **27**, 1505–1511 (1999).
136. Seib, K. L., Peak, I. R. A. & Jennings, M. P. Phase variable restriction-modification systems in *Moraxella catarrhalis*. *FEMS Immunol. Med. Microbiol.* **32**, 159–165 (2002).
137. Zaleski, P., Wojciechowski, M. & Piekarczyk, A. The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. *Microbiology* **151**, 3361–3369 (2005).
138. Fox, K. L. et al. *Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. *Nucleic Acids Res.* **35**, 5242–5252 (2007).
139. Gawthorne, J. A., Beatson, S. A., Srihanta, Y. N., Fox, K. L. & Jennings, M. P. Origin of the diversity in DNA recognition domains in phasevarion associated *modA* genes of pathogenic *Neisseria* and *Haemophilus influenzae*. *PLOS ONE* **7**, e32337 (2012).
140. De Vries, N. et al. Transcriptional phase variation of a type III restriction-modification system in *Helicobacter pylori*. *J. Bacteriol.* **184**, 6615–6623 (2002).
141. Skoglund, A. et al. Functional analysis of the M. HpyAIV DNA methyltransferase of *Helicobacter pylori*. *J. Bacteriol.* **189**, 8914–8921 (2007).
142. Srihanta, Y. N. et al. Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. *PLOS ONE* **6**, e27569 (2011).
143. Srihanta, Y. N. et al. Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. *PLOS Pathog.* **5**, e1000400 (2009).
144. Jen, F. E. C., Seib, K. L. & Jennings, M. P. Phasevarions mediate epigenetic regulation of antimicrobial susceptibility in *Neisseria meningitidis*. *Antimicrob. Agents Chemother.* **58**, 4219–4221 (2014).
145. Srihanta, Y. N. et al. Methylation and phenotypic analysis of the ModH5 phasevarion of *Helicobacter pylori*. *Sci. Rep.* **7**, 16140 (2017).
146. Heithoff, D. M., Sinsheimer, R. L., Low, D. A. & Mahan, M. J. An essential role for DNA adenine methylation in bacterial virulence. *Science* **284**, 967–970 (1999).
147. Garcia-Del Portillo, F., Pucciarelli, M. G. & Casadesús, J. DNA adenine methylase mutants of *Salmonella typhimurium* show defects in protein secretion, cell invasion, and M cell cytotoxicity. *Proc. Natl. Acad. Sci. USA* **96**, 11578–11583 (1999).
148. Brockman, K. L. et al. The ModA2 phasevarion of nontypeable *Haemophilus influenzae* regulates resistance to oxidative stress and killing by human neutrophils. *Sci. Rep.* **7**, 3161 (2017).
149. Brockman, K. L. et al. ModA2 phasevarion switching in nontypeable *Haemophilus influenzae* increases the severity of experimental otitis media. *J. Infect. Dis.* **214**, 817–824 (2016).
150. VanWagoner, T. M. et al. The modA10 phasevarion of nontypeable *Haemophilus influenzae* R2866 regulates multiple virulence-associated traits. *Microb. Pathog.* **92**, 60–67 (2016).
151. Polaczek, P., Kwan, K. & Campbell, J. L. GATC motifs may alter the conformation of DNA depending on sequence context and N6-adenine methylation status: possible implications for DNA-protein recognition. *Mol. Gen. Genet.* **258**, 488–493 (1998).
152. Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
153. Diekmann, S. DNA methylation can enhance or induce DNA curvature. *EMBO J.* **6**, 4213–4217 (1987).
154. Luo, G.-Z. & He, C. DNA N6-methyladenine in metazoans: functional epigenetic mark or bystander? *Nat. Struct. Mol. Biol.* **24**, 503–506 (2017).
155. Fu, Y. et al. N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell* **161**, 879–892 (2015). **This article describes one of the first studies to map 6mA events at high resolution and on the genome scale in an eukaryotic genome.**
156. Mondo, S. J. et al. Widespread adenine N6-methylation of active genes in fungi. *Nat. Genet.* **49**, 964–968 (2017).
157. Greer, E. L. et al. DNA methylation on N6-adenine in *C. elegans*. *Cell* **161**, 868–878 (2015).
158. Zhang, G. et al. N6-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906 (2015).
159. Koziol, M. J. et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2016).
160. Wu, T. P. et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
161. Luo, G.-Z. et al. Characterization of eukaryotic DNA N6-methyladenine by a highly sensitive restriction enzyme-assisted sequencing. *Nat. Commun.* **7**, 11301 (2016).
162. Yin, J. C., Krebs, M. P. & Reznikoff, W. S. Effect of dam methylation on Tn5 transposition. *J. Mol. Biol.* **199**, 35–45 (1988).

163. Ngo, T. T. M. et al. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7**, 10813 (2016).
164. Tan, A., Attack, J. M., Jennings, M. P. & Seib, K. L. The capricious nature of bacterial pathogens: phasevarions and vaccine development. *Front. Immunol.* **7**, 586 (2016).
165. Banerjee, S. & Chowdhury, R. An orphan DNA (cytosine-5)-methyltransferase in *Vibrio cholerae*. *Microbiology* **152**, 1055–1062 (2006).

## Acknowledgements

The authors thank A. Tourancheau and other members of the Fang laboratory for their comments. The work was funded by

R01 GM114472 (G.F.) and R01 GM128955 (G.F.) from the National Institutes of Health. G.F. is an Irma T. Hirsch/Monique Weill-Caulier Trust Research Scholar and a Nash Family Research Scholar.

## Author contributions

All authors researched data for the article, made substantial contributions to discussions of the content and reviewed and/or edited the manuscript before submission. J.B. and G.F. wrote the article.

## Competing interests

J.B. is currently employed by Oxford Nanopore Technologies, Ltd. G.F. and E.S. declare no competing interests.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Reviewer information

*Nature Reviews Genetics* thanks J. Casadesus, M. Oggioni and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## RELATED LINKS

REBASE database:

<http://rebase.neb.com/cgi-bin/pacbiolist?0>