Published in final edited form as:

Curr Opin Microbiol. 2013 April ; 16(2): 192–198. doi:10.1016/j.mib.2013.01.011.

Entering the era of bacterial epigenomics with single molecule real time DNA sequencing

Brigid M. Davis, Michael C. Chao, and Matthew K. Waldor

Division of Infectious Diseases, Brigham and Women's Hospital and Department of Microbiology and Immunobiology, Harvard Medical School and HHMI, 181 Longwood Ave., Boston MA 02115

Abstract

DNA modifications, such as methylation, guide numerous critical biological processes, yet epigenetic information has not routinely been collected as part of DNA sequence analyses. Recently, the development of single molecule real time (SMRT) DNA sequencing has enabled detection of modified nucleotides (e.g, 6mA, 4mC, 5mC) in parallel with acquisition of primary sequence data, based on analysis of the kinetics of DNA synthesis reactions. In bacteria, genome-wide mapping of methylated and unmethylated loci is now feasible. This technological advance sets the stage for comprehensive, mechanistic assessment of the effects of bacterial DNA methyltransferases - which are ubiquitous, extremely diverse, and largely uncharacterized – on gene expression, chromosome structure, chromosome replication, and other fundamental biological processes. SMRT sequencing also enables detection of damaged DNA and has the potential to uncover novel DNA modifications.

Introduction

The genetic blueprints for all organisms are contained in their genomes. The order of the four primary nucleotides - A, T, G, and C – specifies the sequences of the RNAs and proteins that mediate cellular functions. Additional 'epigenetic' information is conveyed through chemical modification of the four bases, which has been observed in all kingdoms of life and has profound biological consequences. For example, cytosine methylation is essential for growth and development in higher eukaryotes, largely through its effect on regulation of gene expression; aberrant cytosine methylation has been implicated in a variety of human diseases, including cancer and neuropsychiatric disorders [1]. Derivatives of methylated cytosine, including hydroxymethylcytosine and the recently discovered formylcytosine and 5-carboxyl cytosine [2,3], may likewise regulate gene expression, although potentially by a different mechanism(s) [4]. The importance of DNA methylation is also well-established in several model proteobacteria, including *E. coli* and *Caulobacter crescentus*, in which methylation of adenine residues (by Dam and CcrM, respectively) is pivotal in the control of chromosome replication, DNA repair, and gene expression (reviewed in [5,6]). However, the enzymes that mediate DNA methylation in prokaryotes are far more diverse than in eukaryotes [7], and the nature, extent and consequences of DNA modification have not been extensively investigated for much of the bacterial kingdom.

©2013 Elsevier Ltd. All rights reserved.

#correspondence: mwaldor@rics.bwh.harvard.edu, 617-525-4646.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The advent of new sequencing platforms in the last decade has allowed the pace of whole genome sequencing to increase exponentially. Meanwhile, large scale analyses of DNA modification have lagged far behind, resulting in a widening gap between the extent of genomic and epigenomic knowledge. The absence of high throughput methods to detect DNA modifications with strand-specificity and nucleotide level resolution on a genome wide scale has markedly hindered our capacity to comprehensively analyze the biologic significance of such modification. However, the recent development of single molecule real time (SMRT) DNA sequencing enables simultaneous elaboration of both genomic and epigenomic information from native bacterial DNA [8]*, [9]**. Thus, we are now poised to begin an exciting era of characterizing the roles of DNA modifications throughout the bacterial kingdom. Currently, the capacity of SMRT sequencing is not sufficient for genome-wide analysis in higher eukaryotes, but such analyses are a downstream goal. Before describing SMRT and the possibilities this technology offers, we will briefly review the types and known consequences of DNA modifications found in bacteria.

DNA modification in the bacterial kingdom

The most widespread DNA modification identified in prokaryotes to date is DNA methylation. Methylation is catalyzed by 3 families of DNA methyltransferases (MTases), which covalently attach methyl groups to either adenine or cytosine residues, usually in a sequence-specific manner. Two classes of enzymes act upon exocyclic amino groups; they yield methyladenine (6mA) and 4 methylcytosine (4mC) [10], which are either absent or extremely rare in mammalian DNA [11]. Additionally, bacteria encode a class of MTases that are homologous to eukaryotic MTases and catalyze the formation of 5 methylcytosine (5mC) [12]. There is extensive variability among bacterial species in their number of MTases and the sequence motifs that are modified, even among closely related organisms and enzymes. Nonetheless, genomic analyses suggest that some form of DNA methylation is present in nearly all bacteria, as putative DNA MTases were found in at least 94% of 3300+ sequenced bacterial genomes (<http://tools.neb.com/~vincze/genomes/>). However, the precise sequence targets and biological roles of the vast majority of these MTases remain virtually unknown. The biological roles of other forms of bacterial DNA modification, such as phosphorothioation, have also not been well characterized [13,14].

Genomic analyses have revealed that many bacterial DNA MTases are encoded in the vicinity of restriction endonucleases (REs), suggesting that they are components of restriction-modification (RM) systems. Such pairs of MTases and REs recognize the same target sequences; methylation typically protects the target site from digestion by the RE. RM systems are generally believed to aid in protecting cells from invading foreign DNA (e.g., phage, conjugative elements, exogenous plasmids, etc.), which is likely to lack protective methylation and hence be subject to RE cleavage [15]. However, recent work has found that RM systems may also play important roles in regulating expression of native DNA sequences: in several species, phase variation of RM-associated MTases seems to modulate gene expression, although the precise mechanisms underlying this effect have not been described [16,17]. It has been postulated that such phase variation enables a single strain to adopt multiple distinct phenotypes, which may be beneficial in adapting to diverse environmental niches. Variable expression of RM genes may also allow modulation of bacterial receptiveness to genetic exchange [18].

Regulatory roles have also been described for so-called “orphan” MTases, which occur in the genome without an associated RE. Comprehensive genomic methylation by orphan MTases is not necessary, since the presence of unmethylated sites is not necessarily deleterious. The archetypal orphan MTase, Dam, which is a 6mA MTase found in a variety of enteric bacteria, has been shown to regulate initiation of chromosome replication, DNA

repair, and gene expression, and to modulate the virulence of several pathogens (reviewed in [5,19]. Most Dam-dependent regulation results from the effects of DNA methylation status upon binding of proteins to their target sites. In similar fashion, CcrM, another 6mA orphan MTase, is essential for the replication and viability of *C. crescentus*, where it controls both the timing of the cell cycle and cellular differentiation [20,21]. However, even for these standout examples, comprehensive, genome-wide analyses of methylation and its consequences have not been performed, while the biological significance of the vast majority of orphan MTases, including most eukaryote-like 5mC MTases, remains entirely unexplored. Thus, the study of both orphan and RM-associated MTases is an open frontier for characterizing how these enzymes can coordinate basic bacterial processes.

Methods for Detection of Modified Bases

To date, detection of modified bases has not routinely been a component of sequence analyses, and it has posed significant technical challenges, particularly in the detection of 6mA, which appears to be the most common modification within bacterial genomes. DNA amplification removes epigenetic marks; therefore, detection of modified nucleotides can only be performed using native DNA as template. Furthermore, while it is possible to detect some modified bases with Sanger DNA sequencing [22], the technique has not been widely adopted for this purpose. ‘Next generation’ high throughput DNA sequencing technologies that require amplification of DNA template during the sequencing process (such as offered by the Illumina HiSeq machine) cannot be used for direct detection of modified bases in native DNA; however, treatment of DNA samples with bisulfite, which converts unmodified cytosines to uracil, does enable discrimination between modified and unmodified cytosines using various sequencing platforms. Still, sample preparation for bisulfite sequencing is cumbersome, and this technique has not allowed discrimination between 5mC and 5hmC [23], although methods have recently been developed to circumvent the latter issue [24,25]. Affinity purification techniques, coupled with microarrays or with DNA sequencing, have proven useful for detection of modified bases, but these modification-specific protocols do not allow open-ended assessment of base modifications or precise mapping of their positions and may require extensive in vitro modification of genomic DNA [26]. Finally, modified nucleotides can also be detected by non-sequencing based approaches such as chromatographic techniques and mass spectrometry, but these methods do not routinely yield genome-wide information with nucleotide level resolution [23].

Detection of modified bases with SMRT DNA sequencing

SMRT sequencing, as the acronym suggests, involves single molecule biochemistry where the synthesis of DNA is monitored in real time. In SMRT sequencing, zero-mode waveguides (nano-structures which enable individual molecules to be isolated for optical analysis) are used to isolate single molecules of a phi29-derived DNA polymerase bound to a single, circular molecule of template DNA [27] [28], and synthesis of DNA from these templates is performed using nucleotides with a distinct fluorophore linked to each of the four bases. Nucleotide incorporation, which yields a fluorescent “pulse”, is monitored for thousands of these reactions in parallel in order to generate the primary sequence data. In addition, the width and time between pulses (pulse width (PW) and interpulse duration (IPD), respectively), yield information regarding the kinetics of DNA synthesis (Figure 1). The high coverage that is readily obtained in SMRT sequencing, as well as the circular templates from which both strands of the primary sequence can be read, facilitate robust statistical analyses of the sequence and associated kinetic data.

Flusberg et. al. used synthetic DNA templates to demonstrate that there are characteristic variations (aka ‘kinetic signatures’ or KV) in the IPD and PW associated with distinct

modifications of deoxyribonucleotides [29]**. That is, as the DNA polymerase encounters a modified base in the template DNA (e.g., 6mA, 5mC, 5hmC, and 4mC), synthesis slows down in a manner that is primarily determined by the identity of the modified base, although the local sequence context also has an effect [29]** [30]**. In particular, 6mA and 4mC are associated with reliable, robust kinetic signatures; the effect of 5mC on synthesis kinetics is less marked, probably because the methyl group in 5mC is not directly involved in base pairing [30]**. Damaged bases, including 8-oxoguanine, O6-methylguanine, 5-hydroxymethyluracil and thymine dimers, have also been shown to give rise to distinct kinetic signatures during SMRT sequencing of synthetic templates [31]. Typically, the effects of modification on synthesis kinetics have been quantified by comparing the average IPDs for native (modified) and amplified (unmodified) templates at each site, and identifying sites where the ratio of the IPDs from the two samples differs from the baseline (Figure 1). As an alternative to the IPD ratio, the distributions of IPDs at each site can be statistically analyzed in native and amplified samples [9]**. Variance in polymerase kinetics is usually most pronounced directly opposite the modified base in the template strand, but it can occur throughout the entire ~11 base region contacted by the polymerase [23], possibly due to long-range distortion of the DNA structure by the presence of a modified base [32]. Recently, Schadt and colleagues developed a sophisticated statistical framework for analyses of kinetic variation (KV) in SMRT sequence data that allows for incorporation of kinetic data from neighboring sites as well as the primary site of interest, in order to more confidently detect modified loci [33]. This strategy may facilitate comprehensive detection of more subtle kinetic variations, such as those induced by 5mC; however, the model has yet to be applied to genome-wide datasets.

Recent applications of SMRT-based detection of methylation in bacteria

SMRT-based sequencing has already proven highly useful for the rapid and simple determination of the specificity of previously uncharacterized MTases. Clarke and colleagues introduced plasmids enabling expression of 16 distinct MTases of both known and unknown specificities into a strain of *E. coli* that was engineered to lack any MTases [30]**. Subsequently, the plasmids were sequenced, and the kinetics of DNA synthesis using modified vs. unmodified plasmid templates were compared. From these analyses, Clarke et. al. could identify the nucleotide targets and sequence specificities of MTases that generate either 6mA or 4mC, even for MTases that are part of Type III RM systems, which have been hard to study with other methods. MTases that create 5mC yielded more subtle kinetic signatures, though in most cases it was possible to distinguish modified from unmodified cytosines. Interestingly, “off-target” signals were observed for a few of the MTases tested, raising the possibility that MTase fidelity may not be as strict as that of restriction enzymes; however, subsequent studies ([8]*[9]**; discussed below) suggest that at least some of the off-target sites detected using this plasmid-based system may have resulted from overexpression of the MTases.

Two recent studies convincingly demonstrate the potential of SMRT sequencing to comprehensively catalog genome-wide DNA methylation in bacteria [8]*[9]**. The study by Murray and colleagues [8]* evaluated the extent of 6mA and 4mC methylation in the genomes of 6 different bacteria, each of which was predicted to encode only a few MTases. Fang et. al. [9]** evaluated 6mA methylation in a Shiga-toxin producing *E. coli* O104:H4 strain isolated during the 2011 HUS outbreak in Germany [34], a strain that was predicted to encode 10 MTases that generate 6mA. Both studies used informatics to deduce motifs that were highly enriched for KV; individual motifs were then shown to be modified by specific MTases using the plasmid-based system mentioned above. New MTase specificities were also discovered, including an enzyme that modifies bases on only one strand of DNA [9]**, whose biological significance warrants further investigation. Additionally, genes encoding

new promiscuous (lacking sequence specificity) MTases were identified in both studies; these genes, like the recently described *hin1523*, *nma1821*, and *hia5* [35], which encode unrelated non-specific MTases, were phage encoded. It is reasonable to speculate that nonspecific methylation of phage DNA may constitute a defense mechanism that facilitates infection of new hosts, which may produce a wide variety of REs. Activation of such methylation may be linked to specific stimuli, since the nonspecific MTases were not found to be active in their native hosts under the conditions tested [9]** [35]. Interestingly, relatively little off-target site methylation was detected for MTases that are part of RM systems, despite the fact that there is no obvious biologic imperative for maintaining exquisite specificity of these enzymes as long as methylation occurs for all cognate RE target sites [8]*, [9]**. For almost all of the adenine MTases other than the nonspecific MTases, greater than 90% of the deduced target sites were found to be methylated in the native hosts, and only a very small fraction of sites were confidently claimed to be unmethylated. In contrast, less than half of the motifs modified by the two 4mC enzymes studied, M1.BceSIII and M2.BceSIII, appeared to be methylated [8]*; however, this may reflect the lower intensity of the 4mC vs the 6mA signal in SMRT data, particularly since these two enzymes are constituents of an RM system, which typically mandates comprehensive methylation. Collectively, these two studies suggest that SMRT will be extremely valuable tool for discovery of novel MTase activities, some of which will likely prove to be useful reagents for molecular biology.

In addition to enabling comprehensive genomic mapping of methylated sites, SMRT sequencing allows quantitation of the frequency with which a particular genomic locus is modified. The study by Fang et. al. [9]** identified a significant number of Dam sites that were methylated in only a fraction of the DNA analyzed. Such heterogeneous methylation might reflect the presence of cells at different stages of the cell cycle; alternatively, it might reflect stable maintenance of distinct methylation states at these loci, or simply stochastic variability. SMRT sequencing can also identify MTase target sites that are uniformly unmethylated in a population, although such analysis requires greater sequence depth than does identification of modified loci. Unmethylated targets, which were found both within the context of hemimethylated sequences (i.e., one of two strands methylated) and in fully unmethylated sequences, likely reflect binding of these sites by proteins that block MTase activity. Consequently, comparative analyses of methylation may yield insight into different DNA-binding activities associated with particular growth conditions or mutations.

Using epigenomic information to address biologic questions

Epigenomic profiling is likely to be particularly powerful when coupled with additional investigative approaches, such as transcriptome analyses. For example, Fang et. al. [9]** compared methylation and gene expression profiles of wildtype Shiga toxin-producing *E. coli* O104:H4 and a mutant strain lacking a phage-borne functional RM system. They identified the target sites for the MTase, as described above, as well as extensive changes in the bacterial transcriptome in the absence of the MTase: expression of nearly 40% of this pathogen's genes was altered [9]**. It is not clear why the effects of this horizontally transmitted RM-encoded MTase on gene expression appear to be far more dramatic than those linked to phase-variable RM systems; it could reflect either differences in the underlying biology or the use of different approaches for assaying gene expression. The frequency of methylation sites near or within genes regulated by this MTase was only enriched for a subset of the gene families whose expression was altered in the MTase mutant, suggesting that in many cases methylation affects gene expression indirectly (e.g. by modulating expression of regulatory factors or of chromosome structure). Unexpectedly, the SMRT sequencing analysis also revealed that fragments of three phages were amplified in the MTase mutant, suggesting that this phage-encoded MTase (M.EcoGIII) also regulates

the replication of unlinked phages in the *E. coli* O104:H4 genome. Finally, the study identified reproducibly unmethylated and hemimethylated sites, which enables prediction and monitoring of loci likely to be protected by regulatory DNA binding proteins. Collectively, the wealth of observations that resulted from this initial study of a bacteria's methylome strongly suggests that there is much to be learned regarding the roles of DNA methylation (and presumably other DNA modifications as well) in the bacterial kingdom.

Many intriguing questions regarding the role of genome modification in bacteria will be easier to address using SMRT-based sequencing, since genome-wide data regarding the sites and extents of methylation and how they are affected by genetic and environmental changes can be correlated with many other cellular attributes. In particular, further investigation is warranted of the relationship between methylation and gene expression, and the means by which one can affect the other (e.g., effects on transcription elongation and chromosome structure, as well as on transcription initiation and gene silencing). The extent of variability in methylation, particularly that produced by orphan MTases or as off-target modifications by RM-encoded MTases, both of which are not linked to selective pressure against the absence of methylation, is also almost entirely unknown. SMRT sequencing may also yield greater understanding of the process of RM-mediated phase variation and of the processes by which cells adapt to the presence of new (e.g., horizontally transmitted) RM systems. SMRT sequencing can also provide an indication of whether RM systems encode functional and expressed RE in the absence of direct biochemical characterization: significantly less than comprehensive methylation across the genome by an RM-associated MTase suggests the cognate RE is unlikely to be active. In such cases, MTases that were acquired as part of RM systems may have been co-opted by the host for regulatory purposes. Finally, analyses of genome-wide DNA modification are likely to facilitate a greater understanding of the extent and nature of DNA damage, both in response to damaging agents and under normal growth conditions.

Conclusion: future applications and challenges

With improvements in SMRT sequencing (e.g., more sensitive detection of 5mC) and the continued development of emerging technologies with similar capacities to produce epigenomic data (e.g., nanopore-based sequencing [36]), it should be possible to comprehensively catalogue all base modifications in bacterial genomes. Furthermore, as the throughput of SMRT sequencing increases, similar studies in higher eukaryotes will become possible. At present, SMRT could be used to gather epigenomic information in mitochondria and simpler eukaryotes (with small genomes), some of which are already known to have unusual and biologically significant chemical modifications of their DNA (e.g., base J in kinetoplastid genomes [37,38]). Furthermore, it seems likely that emerging epigenomic technologies will lead to discovery of additional biologically significant base modifications; new modifications (e.g., 5-formylcytosine and 5-carboxycytosine) have been discovered as recently as 2011 [3] [2]. Moreover, these technologies should enable fundamental new insights into how epigenetic modifications modulate the expression of genomic information. Their effects on DNA binding by regulatory proteins, transcription initiation, elongation, and processivity, and on chromosome structure are all worthy of significant investigation. Additionally, epigenomic information may facilitate assembly of individual genomes from metagenomic data, as proposed by Clark [30]. Finally, epigenomic data may have applications beyond biology. For example, if the patterns of DNA modifications vary with environmental conditions, epigenomic data may prove invaluable for forensic analyses.

Acknowledgments

Grants from the Howard Hughes Medical Institute and NIH-AI-R37 42347 to MKW supported this work.

References

- Ooi SKT, O'Donnell AH, Bestor TH. Mammalian cytosine methylation at a glance. *Journal of Cell Science*. 2009; 122:2787–2791. [PubMed: 19657014]
- He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011; 333:1303–1307. [PubMed: 21817016]
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 2011; 333:1300–1303. [PubMed: 21778364]
- Kellinger MW, Song C-X, Chong J, Lu X-Y, He C, Wang D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol*. 2012; 19:831–833. [PubMed: 22820989]
- Casadesus J, Low D. Epigenetic Gene Regulation in the Bacterial World. *Microbiology and Molecular Biology Reviews*. 2006; 70:830–856. [PubMed: 16959970]
- Low DA, Casadesús J. Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol*. 2008; 11:106–112. [PubMed: 18396448]
- Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*. 2009; 38:D234–D236. [PubMed: 19846593]
- Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. The methylomes of six bacteria. *Nucleic Acids Res*. 2012 Used SMRT DNA sequencing to identify targets for known and uncharacterized DNA methyltransferases in several bacterial species that encode multiple methyltransferases
- Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, et al. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol*. 2012 Used SMRT DNA sequencing to comprehensively identify genomic methylated loci in *E. coli* O104:H4 and to determine the sequence targets for 10 uncharacterized DNA methyltransferases. The effect of deleting a methyltransferase gene upon gene expression and other biological process was also assessed.
- Timinskas A, Butkus V, Janulaitis A. Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene*. 1995; 157:3–11. [PubMed: 7607512]
- Ratel D, Ravanat J-L, Charles M-P, Platet N, Breuillaud L, Lunardi J, Berger F, Wion D. Undetectable levels of N6-methyl adenine in mouse DNA: Cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. *FEBS Lett*. 2006; 580:3179–3184. [PubMed: 16684535]
- Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res*. 1994; 22:1–10. [PubMed: 8127644]
- Xie X, Liang J, Pu T, Xu F, Yao F, Yang Y, Zhao YL, You D, Zhou X, Deng Z, et al. Phosphorothioate DNA as an antioxidant in bacteria. *Nucleic Acids Res*. 2012; 40:9115–9124. [PubMed: 22772986]
- Wang L, Chen S, Vergin KL, Giovannoni SJ, Chan SW, DeMott MS, Taghizadeh K, Cordero OX, Cutler M, Timberlake S, et al. DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proceedings of the National Academy of Sciences*. 2011; 108:2963–2968.
- Boyer HW. DNA restriction and modification mechanisms in bacteria. *Annu. Rev. Microbiol*. 1971; 25:153–176. [PubMed: 4949033]
- Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, Kwok T, Grimmond SM, Robins-Browne RM, Jennings MP. Phasevarion Mediated Epigenetic Gene Regulation in *Helicobacter pylori*. *PLoS ONE*. 2011; 6:e27569. [PubMed: 22162751]
- Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc. Natl. Acad. Sci. U.S.A.* 2005; 102:5547–5551. [PubMed: 15802471]

18. Srikhanta YN, Fox KL, Jennings MP. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat Rev Micro.* 2010; 8:196–206.
19. Wion D, Casadesús J. N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. *Nat Rev Micro.* 2006; 4:183–192.
20. Collier J, McAdams HH, Shapiro L. A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 2007; 104:17111–17116. [PubMed: 17942674]
21. Collier J. Epigenetic regulation of the bacterial cell cycle. *Curr. Opin. Microbiol.* 2009; 12:722–729. [PubMed: 19783470]
22. Bart A, van Passel MWJ, van Amsterdam K, van der Ende A. Direct detection of methylation in genomic DNA. *Nucleic Acids Res.* 2005; 33:e124. [PubMed: 16091626]
23. Korlach J, Turner SW. Going beyond five bases in DNA sequencing. *Current Opinion in Structural Biology.* 2012; 22:251–261. [PubMed: 22575758]
24. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, Balasubramanian S. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science.* 2012; 336:934–937. [PubMed: 22539555]
25. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, et al. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell.* 2012; 149:1368–1380. [PubMed: 22608086]
26. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 2011; 473:394–397. [PubMed: 21552279]
27. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science.* 2003; 299:682–686. [PubMed: 12560545]
28. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009; 323:133–138. [PubMed: 19023044]
29. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Meth.* 2010; 7:461–465. Demonstrated that DNA modification results in reproducible effects upon the kinetics of in vitro DNA synthesis, enabling simultaneous detection of primary sequence and template methylation status
30. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* 2011 Used a plasmid-based system to identify the specific motifs targeted by individual methyltransferases, as well as the precise modifications that they produced
31. Clark TA, Spittle KE, Turner SW, Korlach J. Direct detection and sequencing of damaged DNA bases. *Genome Integr.* 2011; 2:10. [PubMed: 22185597]
32. Johnson SJ, Beese LS. Structures of mismatch replication errors observed in a DNA polymerase. *Cell.* 2004; 116:803–816. [PubMed: 15035983]
33. Schadt E, Banerjee O, Fang G, Feng Z, Wong W, Zhang X, Kislyuk A, Clark T, Luong K, Keren-Paz A, et al. Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Research.* 2012
34. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin C-S, Iliopoulos D, et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N. Engl. J Med.* 2011; 365:709–717. [PubMed: 21793740]
35. Drozd M, Piekarowicz A, Bujnicki JM, Radlinska M. Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Res.* 2012; 40:2119–2130. [PubMed: 22102579]
36. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* 2012; 30:349–353. [PubMed: 22446694]

37. van Luenen HGAM, Farris C, Jan S, Genest P-A, Tripathi P, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramasamy G, et al. Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in Leishmania. *Cell*. 2012; 150:909–921. [PubMed: 22939620]
38. Borst P, Sabatini R. Base J: Discovery, Biosynthesis, and Possible Functions. *Annu. Rev. Microbiol.* 2008; 62:235–251. [PubMed: 18729733]
39. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37:W202–W208. [PubMed: 19458158]

Highlights

SMRT DNA sequencing allows high-resolution detection of modified nucleotides

Methylated loci in bacterial genomes can be comprehensively and quantitatively mapped

SMRT DNA analyses reveal specific sequences targeted by methyltransferases

Methylation status can be correlated with numerous biological processes

Roles for DNA methylation and methyltransferases are uncharacterized in most bacteria

Figure 1.

Pipeline for SMRT-based analyses of DNA methylation and MTases

(A) For genome wide analyses, native, methylated (star) genomic and whole genome amplified (unmodified) DNA are used as templates for SMRT sequencing. To analyze the target specificities of individual MTases, amplified and native plasmid DNA from a strain expressing a single MTase can be sequenced.

(B) At a methylated residue, the DNA polymerase (DNAP) pauses before inserting a fluorescently labeled dNTP into the newly synthesized DNA.

(C) The delay in dNTP incorporation opposite a modified base is seen on a chromatogram as an extended interpulse duration (IPD).

(D) Calculation of the IPD ratio for native and unmodified DNA at each position reveals kinetic variation (KV) that exceeds baseline levels at the site of modification. Statistical analyses can be used to identify sites that can confidently be called as modified. Neighboring sites may also display altered kinetics (not shown) that can enhance these analyses.

(E) The sequence contexts of sites showing KV can be analyzed (e.g., using MEME [39]) in order to identify consensus sequence motifs that are targeted by MTases.

(F) After generation of a genome-wide profile of DNA methylation, specific MTases can be functionally characterized, e.g., by comparing the transcriptomes of MTase-deficient mutants and to wildtype cells.

(F) Additional analyses that may illuminate the biological roles of MTases and DNA modification.

Figure 2.
Chemical structures of common modified bases generated by DNA methyltransferases.