

# Exploring the roles of DNA methylation in the metal-reducing bacterium *Shewanella oneidensis* MR-1

Matthew L. Bendall <sup>1</sup>, Khai Luong <sup>2</sup>, Kelly M. Wetmore <sup>3</sup>, Matthew Blow <sup>1</sup>, Jonas Korlach <sup>2</sup>, Adam Deutschbauer <sup>3</sup>, and Rex R. Malmstrom <sup>1#</sup>

1) DOE Joint Genome Institute, Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

2) Pacific Biosciences, Menlo Park, CA, USA

3) Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>1#</sup>*To whom correspondence may be addressed. E-mail: rrmalmstrom@lbl.gov*

August 26, 2013

## ACKNOWLEDGMENTS:

*The work conducted by the US Department of Energy (DOE) Joint Genome Institute is supported by the Office of Science of the DOE under Contract Number DE-AC02-05CH11231.*

## DISCLAIMER:

**[LBNL]** This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California

Exploring the roles of DNA methylation in the metal-reducing bacterium  
*Shewanella oneidensis* MR-1

Matthew L. Bendall<sup>1</sup>, Khai Luong<sup>2</sup>, Kelly M. Wetmore<sup>3</sup>, Matthew Blow<sup>1</sup>, Jonas  
Korlach<sup>2</sup>, Adam Deutschbauer<sup>3</sup>, and Rex R. Malmstrom<sup>1#</sup>

1) DOE Joint Genome Institute, Genomics Division, Lawrence Berkeley National  
Laboratory, Berkeley, CA, USA

2) Pacific Biosciences, Menlo Park, CA, USA

3) Physical Biosciences Division, Lawrence Berkeley National Laboratory,  
Berkeley, CA, USA

# Corresponding author: rrmalmstrom@lbl.gov

**ABSTRACT**

We performed whole genome analyses of DNA methylation in *Shewanella oneidensis* MR-1 to examine its possible role in regulating gene expression and other cellular processes. Single-Molecule Real Time (SMRT) sequencing revealed extensive methylation of adenine (N6mA) throughout the genome. These methylated bases were located in five sequence motifs, including three novel targets for Type I restriction/modification enzymes. The sequence motifs targeted by putative methyltransferases were determined via SMRT sequencing of gene knockout mutants. In addition, we found *S. oneidensis* MR-1 cultures grown under various culture conditions displayed different DNA methylation patterns. However, the small number of differentially methylated sites could not be directly linked to the much larger number of differentially expressed genes in these conditions, suggesting DNA methylation is not a major regulator of gene expression in *S. oneidensis* MR-1. The enrichment of methylated GATC motifs in the origin of replication indicate DNA methylation may regulate genome replication in a manner similar to that seen in *Escherichia coli*. Furthermore, comparative analyses suggest that many *Gammaproteobacteria*, including all members of the *Shewanellaceae* family, may also utilize DNA methylation to regulate genome replication.

## 47 INTRODUCTION

48 DNA methylation plays a variety of functional roles in bacteria [1-3]. For  
49 example, restriction-modification (R-M) systems use methylation patterns to  
50 identify and destroy foreign DNA during viral infections [4, 5]. Bacteria also use  
51 DNA methylation to regulate genome replication [6], DNA mismatch repair [7],  
52 and gene expression [8-12]. Methylation can even serve as an epigenetic  
53 modifier, influencing the expression patterns of daughter cells based on  
54 environmental conditions [13, 14]. Because of these varied regulatory roles,  
55 DNA methylation should be incorporated into our emerging systems-level view of  
56 model microorganisms.

57 Despite the functional significance of DNA methylation, our understanding  
58 of its role in bacterial genetics and physiology remains incomplete due to  
59 methodological limitations. For example, bisulfite conversion can identify 5-  
60 methylcytosine modifications [15, 16], but there is no corresponding conversion  
61 assay for other common modifications in bacteria such as N6-methyladenine or  
62 4-methylcytosine [17]. Methyl-sensitive restriction enzymes have been used to  
63 identify the methylation state of specific sequence motifs [18-20], but complete  
64 methylome analyses are not possible without 1) prior knowledge of the entire set  
65 of methyltransferases and their sequence targets within a genome, and 2)  
66 access to methyl-sensitive restriction enzymes targeting these motifs. Single-  
67 Molecule Real Time (SMRT) sequencing overcomes these limitations and  
68 enables genome-wide analysis of DNA methylation with single base resolution  
69 [21]. In this approach, modifications in the native state DNA are revealed by

70 deviations in the polymerase kinetics observed during sequencing. The specific  
71 type of DNA methylation can often be determined from the polymerase kinetics,  
72 e.g. N6-methyladenine or 4-methylcytosine. With SMRT sequencing it is now  
73 possible to identify the complete set of methylated sequence motifs within a  
74 microbial genome as well as the methylation state for each instance of a motif  
75 [22-24]. This represents a powerful tool for characterizing the functional roles of  
76 DNA methylation in a wide variety of bacteria.

77       *Shewanella oneidensis* MR-1 is a bacterial isolate belonging to the  
78 *Shewanellaceae*, a family distinguished by the wide variety of electron acceptors  
79 they can utilize (e.g. iron, manganese, uranium, chromium, and plutonium) [25-  
80 28]. Because of their flexible respiratory pathways, *Shewanella* sp. are  
81 recognized as potential agents for bioremediation at sites contaminated with  
82 heavy metals and radionuclides [29]. To better exploit its metabolic potential, *S.*  
83 *oneidensis* MR-1 has been characterized extensively, including analysis of gene  
84 expression [30, 31], identification of regulatory regions [32], and the  
85 determination of fitness levels for thousands of gene knockout mutants  
86 [33]. However, the developing systems-level view of *Shewanella* does not yet  
87 incorporate DNA methylation and its potential regulatory roles. Genomic  
88 analyses reveal multiple putative methyltransferases in *S. oneidensis* MR-1 [34,  
89 35], including several apparent 'orphans' that lack corresponding restriction  
90 enzymes. It remains unclear what role these orphan methyltransferases might  
91 play.

92           Here we use SMRT sequencing to provide the first look at DNA  
93   methylation in *S. oneidensis* MR-1. We identify methylated sites throughout the  
94   genome as well as the sequence motifs targeted by predicted  
95   methyltransferases. To determine if DNA methylation regulates gene  
96   expression, we examine whether changes in expression level correspond with  
97   changes in DNA methylation state when cultures are transferred from one set of  
98   growth conditions to another. Finally, we examine the finished genomes of all  
99   *Gammaproteobacteria*, including the *Shewanellaceae*, to determine which groups  
100   appear to use DNA methylation for regulating genome replication and DNA  
101   mismatch repair.

102

## 103   **RESULTS AND DISCUSSION**

### 104   *Methylation profile of S. oneidensis MR-1*

105           To identify methylated sites within the genome of *S. oneidensis* MR-1, we  
106   performed SMRT sequencing on DNA extracted from triplicate exponential-phase  
107   cultures grown aerobically on minimal media. Our analysis revealed 42,965  
108   nucleotides that exhibited significant variations in polymerase kinetics that were  
109   diagnostic of DNA modification [21]. Of those modified nucleotides, 41,853 were  
110   identified as N6-methyladenine (N6mA) based on their distinct kinetic fingerprint.  
111   The remaining kinetic variants included 396 cytosine, 301 guanine, and 415  
112   thymine bases (Supplementary Table 1). While some of these were likely  
113   analytical artifacts, the agreement among all three biological replicates suggests  
114   additional, unidentified mechanisms for DNA modification at work in *S.*

115 *oneidensis* MR-1, e.g. glucosylation, putrescinylation, and glutamylation [36,  
 116 37]. The nature of these putative modifications could not be determined in this  
 117 study, but represent an interesting avenue for future investigation.

118       Next we analyzed the local sequence context of these N6mA bases to  
 119 determine if they were located within specific sequence motifs. We identified two  
 120 palindromic motifs (GATC and ATCGAT) and three bipartite motifs (GCAN<sub>4</sub>GTC /  
 121 GACN<sub>4</sub>TGC, TACN<sub>6</sub>GTNGT / ACNACN<sub>6</sub>GTA, and TGAYN<sub>6</sub>TGAC /  
 122 GTCAN<sub>6</sub>RTCA) enriched with N6mA bases (methylated bases are underlined).  
 123 Over 99% of all N6mA residues could be assigned to one of these five sequence  
 124 motifs, suggesting that methylation was performed by DNA methyltransferases  
 125 targeting specific DNA sequences. Genome annotations of *S. oneidensis* MR-1  
 126 predicted three Type I RM systems with unknown target sequences [34]. The  
 127 discovery of three bipartite motifs is consistent with Type I R-M systems [38], but  
 128 the specific motif sequences we detected were novel R-M targets. The two  
 129 methylated palindromes GATC and ATCGAT were also predicted in restriction  
 130 enzyme database REBASE (<http://rebase.neb.com/rebase/rebase.html>) based  
 131 on sequence homology to methylases with known targets [35]. However, *S.*  
 132 *oneidensis* MR-1 has four putative Type II methyltransferases, which is more  
 133 than the number of methylated motifs detected once the three Type-I-like  
 134 bipartite motifs were excluded. All putative methyltransferases were expressed  
 135 under these growth conditions, suggesting either redundancy among the  
 136 methyltransferases in targeted motifs or possible misannotation of some genes.

137 To identify the specific target of each methyltransferase, we performed  
138 SMRT sequencing on methyltransferase transposon mutants. In these tests, a  
139 target motif was assigned to a predicted methyltransferase if the motif was not  
140 methylated in the transposon mutant. For example, the sequence motif ATCGAT  
141 was no longer methylated in a mutant lacking putative methyltransferase  
142 SOA0004, thus indicating the target of this methyltransferase. The three bipartite  
143 motifs were also clearly assigned to the three predicted Type I  
144 methyltransferases using this approach (Table 2). However, the protein(s)  
145 targeting GATC could not be identified unequivocally because this motif was  
146 always methylated in the individual transposon mutants. Interestingly, we were  
147 unable to generate a viable transposon mutant for predicted methyltransferase  
148 SO\_0289, suggesting this gene maybe required for viability. Two of the putative  
149 methyltransferases show homology to *dam* (SO\_0289 & SO\_0690) and were  
150 suspected to target GATC (Table 2). Gene SO\_3004 does not show strong  
151 homology to *dam*, but was also predicted in REBASE to target GATC. Thus, it  
152 appears that *S. oneidensis* MR-1 may use multiple genes to methylate  
153 GATC. Similar redundancies have been observed in *E. coli*, which has three  
154 separate enzymes that methylate GATC [3, 22]. In addition, some strains of  
155 *Salmonella enterica* may have up to five copies of *dam* genes (Supplementary  
156 Table 3). The potential redundancy of GATC-targeting methyltransferases and  
157 the lack of corresponding restriction enzymes suggest that *S. oneidensis* MR-1  
158 may use GATC methylation in some regulatory capacity.

To examine the efficiency of methyltransferases in *S. oneidensis* MR-1, we determined the methylation state for each instance of the five sequence motifs throughout the genome. After excluding 418 motif locations with low sequence coverage, we found that 99.2% of the remaining 21,118 motif locations were methylated on both strands in all three replicates. The degree of methylation varied by motif from as low as 96.9% to greater than 99.9% (Table 1). Interestingly, 5 sites were confidently hemi-methylated, i.e. methylated on only one strand, in all three replicates, whereas 8 sites were not methylated on either strand (Supplementary Table 2). Five of the non-methylated sites were in Type I R-M motifs, two were located in GATC motifs, and one was found in an ATCGAT motif (Table 1; Supplementary Table 2). Non-methylated and hemi-methylated sites have been detected in other bacteria [8, 20, 22, 23, 39, 40], and these sites are often protected from methyltransferases by DNA binding proteins.

Changes in DNA methylation can impact gene expression levels by altering the binding affinity of regulatory proteins. For example, the binding of leucine responsive protein (Lrp), which regulates expression levels of the *pap* operon in *E. coli*, is controlled by methylation state of two GATC sites in the Lrp binding site [8, 41]. In *S. oneidensis* MR-1, GATC motifs are commonly found in the binding sites for transcription factors Crp and Fnr [32]. Indeed, 64 of the 174 Crp binding sites and 21 of 30 Fnr binding sites have one or more Type II motif (Table 3). This enrichment of GATC motifs in transcription factor binding sites presents an opportunity for methylation to impact gene expression on a large scale. That is, if DNA binding activities of Crp and Fnr are sensitive to the DNA

methylation, then changes in the methylation of GATC sites could potentially influence gene expression in *S. oneidensis* MR-1.

#### *Dynamics of DNA methylation and gene expression*

To explore changes in genome-wide methylation patterns and their possible impact on gene expression, we measured DNA methylation and mRNA levels in exponential-phase *S. oneidensis* MR-1 cultures grown under various conditions. More specifically, triplicate cultures grown aerobically in minimal media were first analyzed by SMRT sequencing and RNAseq, and then analyzed again after transfer to either aerobic rich media or anaerobic minimal media. Not surprisingly, gene expression varied from one condition to another, with 426 genes differentially expressed between minimal and rich media (201 up-regulated, 225 down-regulated), and 99 genes differentially expressed between aerobic and anaerobic conditions (51 up-regulated, 48 down-regulated) (Fig 2; Supplementary Tables 4&5).

We did not observe widespread changes in methylation in cultures growing under different conditions. However, reproducible differences in methylation state were observed at 8 sites (Table 4, Supplementary Table 2). Of these sites, 5 were GATC motifs located within intergenic regions. One of these GATC sites (genome position 4,061,174) was located within the binding site of a ferric uptake regulator protein (Fur) transcription factor (Table 4), although neither gene regulated by this transcription factor was differentially expressed. The methylation state of two other GATC motifs flanking another Fur binding site

(genome positions 3,823,765 and 3,823,792) also differed between minimal and rich media. One of the nine genes (SO\_3667) regulated by this binding site was up regulated when cultures were transferred to rich media (Supplementary Table 4). Finally, a differentially methylated GATC motif was detected 215 bp upstream of *argC* (SO\_0275), a gene involved in arginine synthesis. The expression of *argC* and four other genes in the same operon was repressed when cultures were transferred from minimal media lacking arginine into rich media containing arginine (Supplementary Table 4). The latter two cases were suggestive of a possible connection between DNA methylation and gene expression, although the nature of the connection remains unclear. That is, did changes in methylation state influence expression levels, or did binding of regulatory proteins inadvertently protect these sites from methylation? Establishing a direct causal link was not possible with the current dataset and will require additional future investigations.

While these methylation dynamics are intriguing, it is clear that widespread changes in gene expression were not accompanied by equally widespread changes in DNA methylation (Figure 2). Nor was there an obvious relationship between methylation at known transcription factor binding sites and expression. For example, 69 Crp-regulated genes were differentially expressed when cultures were transferred from minimal media to rich media, but none of the Crp binding sites were differentially methylated (Table 3, Supplementary Tables 4 and 5). The only differentially methylated transcription factor binding site we observed did not exhibit a significant difference in expression for the

corresponding genes. It is worth noting that *Shewanella* cultures were not synchronized, and it is conceivable that short-term changes in methylation state and gene expression within a small subset of the culture could have gone undetected. Still, while methylation may impact the expression of a few genes, our data suggest that methylation does not play a large and direct role in regulating gene expression in *S. oneidensis* MR-1, at least not under the conditions we tested.

To our knowledge this is the first study to directly measure the genome-wide dynamics of methylation and gene expression in a wild-type bacterium. Three previous studies examined gene expression in *E. coli* mutants lacking *dam* (Dam-), and in each case the expression patterns of the mutants differed from those of the wild-type strain [42-44]. However, support for a direct regulatory role for methylation was inconsistent among studies. For example, Oshima et al. (2002) found that a disproportionately large number of the differentially expressed genes in Dam- mutants were regulated by transcription factors with GATC motifs within their binding sites. In contrast, Lobner-Olesen et al (2003) and Robbins-Menke et al (2005) found that the regulatory regions of differentially expressed genes were not enriched with GATC for both Dam- and Dam-overproducing mutants. While there are clear cases where methylation directly regulates gene expression *E. coli* [45], the latter two studies suggest that most differential gene expression in dam mutants results from indirect effects of variable Dam concentrations on cell physiology, not from direct regulation of expression via methylation. Similarly, >30% of genes in *E. coli* C227-11 were

differentially expressed after deleting RM.EcoGIII methyltransferase, yet there were no clear connections to RM.EcoGIII methylation sites for most of these genes [22]. These studies highlight the challenges of interpreting direct regulatory roles for DNA methylation by examining methyltransferase knockout mutants. Monitoring methylation dynamics in wild-type bacteria, which is now possible using SMRT sequencing, should provide a clearer picture of the regulatory roles of DNA methylation in bacteria.

#### *Roles for DNA methylation in genome replication and DNA mismatch repair*

If DNA methylation is not a major regulator of gene expression in *S. oneidensis* MR-1, then why does its genome encode orphan methyltransferases? One possibility is that DNA methylation plays a critical role in DNA mismatch repair and/or genome replication. For example, *S. oneidensis* MR-1 contains the DNA mismatch repair gene *muthH*, which in *E. coli* nicks the unmethylated strand near hemi-methylated GATC sites when mismatches occur during genome replication. This enables removal of the unmethylated strand and resynthesis of the correct sequence from the methylated template [7]. In addition, the non-uniform distribution of GATC sites throughout the genome suggests *S. oneidensis* MR-1 may use DNA methylation to regulate genome replication in manner similar to *E. coli*. Specifically, the origin of replication (*oriC*) is enriched with GATC sites in both organisms (Figure 3; Supplementary Figure 1), and in *E. coli* the methylation state of these sites plays an important role in regulating genome replication [46-49]. Briefly, genome replication in *E. coli* is

initiated in part by the binding of DnaA to the origin of replication. Shortly after replication begins SeqA binds to newly-formed hemi-methylated GATC sites in the origin, thus preventing additional binding of DnaA and the re-initiation of replication [46, 50]. SeqA also binds to hemi-methylated GATC sites in the promoter region of *dnaA* and reduces transcription of *dnaA* once the replication fork has passed [6, 51], thereby decreasing the chance of replication re-initiation. We hypothesize *S. oneidensis* MR-1 uses similar mechanisms to control DNA replication based on the presence of *dam* and *seqA* in the genome as well as enrichment of GATC sites in the *oriC*.

To determine if DNA methylation regulates genome replication and mismatch repair in *S. oneidensis* MR-1, we attempted to construct *dam*- and *seqA*- mutants. DNA replication and cell division are not synchronized in *dam*- and *seqA*- mutants of *E. coli* [51, 52], and we wished to observe if the same was true for *S. oneidensis* MR-1. However, both *dam* (locus id SO\_0289) and *seqA* appear to be essential for viability as determined by high-throughput transposon mutagenesis and sequencing (TnSEQ) [53] (unpublished data; KMW, MB, AD), so mutants could not be recovered. *Dam* is also essential for fellow *Gammaproteobacteria* *Vibrio cholerae* and *Yersinia pseudotuberculosis*, where it plays a role in regulating genome replication [54, 55]. Interestingly, mutants were viable for the two other *S. oneidensis* MR-1 genes that either showed some homology to *dam* (SO\_0690) or were predicted to methylate GATC by REBASE (SO\_3004). Both genes were expressed in wild type *S. oneidensis* MR-1, indicating that they were functional, but not completely redundant to

SO\_0289. The non-viability of both *dam*- and *seqA*- mutants suggests that methylation of GATC motifs plays a critical functional role in *S. oneidensis* MR-1, mostly likely in regulating genome replication.

Next we asked if other *Shewanella* species might utilize methylation to regulate genome replication or mismatch repair by searching their genomes for *dam*, *seqA*, *mutH*, and GATC enrichment around at the origin or replication. In addition to *S. oneidensis* MR-1, there are 21 finished *Shewanella* genomes isolated from diverse environments including marine and freshwater sediments, hydrothermal vents, and the nidamental gland of a squid. Nineteen isolates encoded mismatch repair protein MutH, whereas all 21 isolates were positive for *dam*, *seqA*, and GATC enrichment around *oriC* (Figure 4, Supplementary Table 3), the same pattern observed in *S. oneidensis* MR-1 and *E. coli*. This suggests that regulation of genome replication through DNA methylation might be a universal feature of the *Shewanellaceae* family.

Looking beyond the *Shewanellaceae*, we searched all finished *Gammaproteobacteria* genomes for the presence of *dam*, *mutH*, *seqA*, and GATC enrichment in *oriC*. Of the 448 finished *Gammaproteobacteria* genomes in the Integrated Microbial Genomes database [56], 331 encoded one or more putative *dam* genes, while 266 encoded *seqA* and 278 encoded *mutH* (Figure 4; Supplementary Figure 2; Supplementary Table 3). Interestingly, virtually all *Gammaproteobacteria* arising after a single evolutionary branch point were positive for *dam*, *mutH*, *seqA*, and GATC enrichment at *oriC* (Figure 4; Supplementary Figure 2). There were 24 exceptions beyond this branch point

that lacked *seqA* and GATC enrichment, and of these, 23 were endosymbionts with massively reduced genomes (Supplementary Table 3). Genome reduction, including loss of regulatory elements, is a common feature of endosymbionts [57], and it appears both *dam* and *seqA* were lost during these reductions. *Glaciacola nitratireducens* FR1064 is not an endosymbiont, but its genome is >1Mbp smaller than the two other sequenced members of the genus *Glaciacola* [58], suggesting it too may have lost *seqA* during genome reduction. While some basal lineages within the *Gammaproteobacteria* encoded *dam* and/or *mthH* genes, none encoded *seqA* or showed enrichment of GATC sites at the origin of replication (Figure 4). Using the limited number of genomes available at the time, Lobner-Olesen et al. (2005) identified a “DamMT clade” within the *Gammaproteobacteria*. Our more comprehensive analysis generally agrees with this earlier report, although it would be more appropriate to discuss a “*seqA*/GATC-*oriC* clade” since *dam* and *mthH* were not exclusive to one clade of *Gammaproteobacteria* (Figure 4).

The development of epigenetic regulation of genome replication appears to be a key evolutionary event within the *Gammaproteobacteria*. The phylogenetic pattern of *dam*, *seqA*, and GATC enrichment at the origin indicates this mechanism for regulating chromosome replication via DNA methylation was acquired and maintained by more recent lineages, not lost by the more basal groups within the clade. Moreover, this mechanism has been maintained throughout the evolution of multiple families, even as these groups have diversified and expanded into a wide range of different aquatic, terrestrial, and

host-associated environments. Dam has even become essential for viability of some *Gammaproteobacteria* such as *S. oneidensis* MR-1 (this study), *Vibrio cholerae*, and *Yersinia pseudotuberculosis* [54]. The maintenance of *dam*, *seqA*, and GATC enrichment at *oriC* throughout their evolutionary history suggests many *Gammaproteobacteria*, regardless of their environment, experience strong selective pressure to synchronize genome replication with cell division, and that more recently evolved lineages do so with DNA methylation. Presumably, the more basal lineages of *Gammaproteobacteria* lacking *seqA* and GATC enrichment at *oriC* use alternative strategies. Interestingly, many of these basal *Gammaproteobacteria* lineages encode *dam*, but its functional role remains a mystery since they do not appear to use methylation for DNA mismatch repair or genome replication. Further analyses with SMRT sequencing will shed additional light on the role of DNA methylation in these bacteria.

## Conclusions

DNA methylation appears to serve a variety of functions in *S. oneidensis* MR-1 including restriction/modification, DNA mismatch repair, and regulation of genome replication. Methylation may also regulate expression of a few genes, but it does not appear to be a major regulator of gene expression. This connection to gene expression is one of the more interesting, but poorly understood, facets of DNA methylation, and it remains to be seen if it is a major regulator of expression in any bacteria. However, with the exception of *E. coli* and *Caulobacter crescentus*, our understanding of DNA methylation is extremely

limited for the vast majority of prokaryotes. The further application of SMRT sequencing will dramatically expand our understanding of DNA methylation in these understudied clades. Large scale surveys of diverse microbial groups using SMRT sequencing will help provide new insights into the scope and variety of DNA methylation in various phylogenetic groups, whereas analyses of wild-type and knockout mutants will help uncover and experimentally verify the functional roles methylation.

## **METHODS AND MATERIALS**

### *Strains, culture conditions, and nucleic acid isolation*

*S. oneidensis* MR-1 was obtained from the American Type Culture Collection (catalog number 700550). The aerobic minimal media contained per liter: 1.5 g NH<sub>4</sub>Cl, 0.1 g KCl, 1.75 g NaCl, 0.61 g MgCl<sub>2</sub>·6H<sub>2</sub>O, 0.6 g NaH<sub>2</sub>PO<sub>4</sub>, 30 mM PIPES buffer, 20 mM DL-lactate, Wolfe's vitamins, and Wolfe's minerals (pH7). Fumarate was added as an electron acceptor to anaerobic minimal media (30mM final conc.). Anaerobic minimal media experiments were set up in an anaerobic chamber (Coy) with a gas mix of 5% H<sub>2</sub>, 10% CO<sub>2</sub>, and 85% N<sub>2</sub>, and incubated in serum bottles closed with butyl rubber stopper at 200 rpm. Aerobic minimal media experiments were conducted in 10 mL culture tubes or 250 ml culture flasks at 200 rpm. All cultures were incubated at 30°C.

To conduct the initial survey of DNA methylation, wild-type cells were collected during exponential growth in minimal media (OD<sub>600</sub> = 0.32). DNA and RNA were isolated with the DNeasy Blood & Tissue kit (Qiagen) and RNeasy

mini kit (Qiagen), respectively. Cells were treated with RNAprotect (Qiagen) according the manufacturers instructions prior to RNA extraction, and total RNA was isolated after an on-column DNase treatment. To examine changes in methylation and gene expression in wild-type cells, we inoculated exponentially growing cells from aerobic minimal media (OD600 = 0.32) into either rich media (Luria-Bertani broth) or anaerobic minimal media. DNA and RNA were collected after cultures reached exponential phase in the new media (OD600 of 0.45 for rich media, ~5 population doublings post transfer; OD600 of 0.16 for anaerobic media, ~3 population doublings post transfer).

DNA methylation patterns were also determined in mutant strains whose putative methyltransferases were inactivated by transposon insertions [33]. DNA from transposon mutants was collected from stationary phase cultures growing in aerobic minimal media (OD600 of ~0.80). All mutants grew to the same density and did not display any substantial growth differences from wild-type.

#### *DNA methylation detection*

Libraries of replicate of wild-type *S. oneidensis* MR-1 cultures were prepared for SMRT sequencing using a library construction protocol described previously [59]. These libraries were sequenced to a mean genome coverage depth of 118-222X on the Pacific Biosciences RS instrument using C2 chemistry. One replicate grown in anaerobic minimal media was excluded from further analysis due to poor coverage (<70X). Methyltransferase transposon mutants were sequenced to a depth 72-113X. Average read lengths ranged from 2,296-

2,889bp in all libraries. Reads were mapped to the reference genome (RefSeq NC\_004347.1) using BLASR [60]. Base modification and motif detection were performed using the Modification and Motif Detection protocol in SMRTPipe v.1.3.3. Positions with coverage >25X and kinetic scores  $\geq$  QV 40 were considered modified. The kinetic score is the  $-10 \cdot \log(p\text{-value})$  where p-value was determined from a t-test between the sample and the in silico model ([http://pacb.com/applications/base\\_modification/index.html](http://pacb.com/applications/base_modification/index.html)). Positions with coverage <25X were excluded from analysis.

#### *Gene expression analyses*

Total RNA collected from biological replicates was sequenced on the Illumina platform to determine gene expression levels of *S. oneidensis* MR-1 grown under the three experimental conditions described above. Ribosomal RNA was removed using the Ribo-Zero Meta-Bacteria kit (Epicentre) prior to creation of a strand-specific RNAseq library [61]. Libraries were sequenced on the Illumina HiSeq 2000 platform. Ten million mRNA reads randomly selected from each library were analyzed for differential expression using DEseq [62], edgeR [63], and RankProd [64] tools. Differentially expressed genes were those confidently identified by all three tests ( $Q < 0.005$ ;  $FDR < 0.01$ ) as having a fold change  $> 2$ .

#### *Detecting dam, seqA, muth, and GATC enrichment at the origin of replication*

We analyzed all 448 finished *Gammaproteobacteria* genomes available in the Joint Genome Institute's Integrated Microbial Genomes database (<http://img.jgi.doe.gov/>) [56] as of March 6, 2013. The number of genes assigned to COGs 0338, 3057, and 3066 was determined in order to quantify the number of *dam*, *seqA*, and *mutH* genes, respectively. The origin of replication was determined by in silico and/or in vivo predictions retrieved from the DoriC database v6.5 (<http://tubic.tju.edu.cn/doric/index.php>) [65, 66]. For the small number of genomes missing from DoriC, the origin was found using Ori-Finder (<http://tubic.tju.edu.cn/Ori-Finder>) [67]. The origin of replication was considered to be enriched in GATC motifs if the observed number of motifs with the origin was significantly greater than would be expected if GATC motifs were uniformly distributed throughout the genome (i.e. the total number of GATC sites divided by genome size). Significance was calculated using a binomial test with a p-value threshold of 0.01.

In four of the genomes examined, we found that the origin annotated in DoriC was not enriched for GATC despite the presence of *dam* and *seqA*. The annotated origins shared the following characteristics: 1) the origin was predicted by in silico analysis, which considers both *dnaA* and *gidA* to be "indicator genes" of the origin; 2) the proposed origin was located upstream of *dnaA*, whereas the predicted origin in other closely related strains was located upstream of *gidA*; and 3) the region upstream of *gidA* was significantly enriched with GATC motifs. This suggested the apparent anomalies were due to misannotation of the origin or replication. The origin was re-defined around *gidA* in these cases.

457

## 458 *Phylogeny of Gammaproteobacteria*

459       16S rRNA sequences from finished *Gammaproteobacteria* genomes were  
460 aligned using MAFFT v6.864b [68] and used to construct a maximum likelihood  
461 phylogeny with RAxML v7.2.6 [69] using the GTRGAMMA model and rapid  
462 bootstrap algorithm (1,000 iterations).

## 463 **ACKNOWLEDGEMENTS**

464       The DOE Joint Genome Institute was supported by the US Department of  
465 Energy Office of Science under contract number DE-AC02-05CH11231.

466

## 467 **FIGURE LEGENDS**

468       Figure 1: Locations of methylated sequence motifs in the chromosome and  
469 megaplasmid of *S. oneidensis* MR-1. Moving from the outer ring to the inner ring,  
470 the methylated motifs are: GATC, ATCGAT, GCAN<sub>4</sub>GTC/GACN<sub>4</sub>TGC (motif 3),  
471 TACN<sub>6</sub>GTNGT/ACNACN<sub>6</sub>GTA (motif 4), and TGAYN<sub>6</sub>TGAC/GTCAN<sub>6</sub>RTCA  
472 (motif 5). The length of each methylation site marker is proportional to the  
473 confidence of methylation; all indicated sites have a p-value < 0.0001. Hemi-  
474 methylated (●) and non-methylated (○) sites in cultures grown aerobically in  
475 minimal media are indicated. OriC indicates the position of the origin of  
476 replication.

477

478 Figure 2: Locations of differentially expressed genes (red and blue rings) and  
479 differentially methylated motifs in *S. oneidensis* MR-1 cultures transferred from  
480 aerobic minimal media to rich media (A) or anaerobic minimal media (B). Up-  
481 regulated genes are red and down-regulated genes are blue. Motif locations that  
482 became methylated (●), hemi-methylated (◐), or non-methylated (○) after  
483 transfer to different culture conditions are indicated. No differentially methylated  
484 motifs were found in the megaplasmid.

485

486 Figure 3: Distribution of GATC motifs throughout the genomes of *S. oneidensis*  
487 MR-1 and *E. coli*. The arrow indicates the origin of replication (*oriC*).

488

489 Figure 4: Phylogenetic patterns of DNA methylation usage in  
490 *Gammaproteobacteria*. The presence of *dam*, *seqA*, and GATC enrichment at  
491 the origin of replication is illustrated on a phylogenetic tree constructed from the  
492 16S sequences of 448 finished genomes. The scale bar represents the fraction  
493 of substitutions per site. The star indicates a node beyond which all but 24  
494 genomes encoded *dam*, *seqA*, and were enriched for GATC motifs in the origin  
495 of replication. Although this node does not have strong bootstrap support based  
496 on 16S rRNA sequences, this node was supported by a rare indel in RpoB [70,  
497 71].

498

499

500

501

## REFERENCES

1. Casadesus J, Low D: **Epigenetic gene regulation in the bacterial world.** *Microbiology and Molecular Biology Reviews* 2006, **70**(3):830-+.
2. Low DA, Casadesus J: **Clocks and switches: bacterial gene regulation by DNA adenine methylation.** *Current Opinion in Microbiology* 2008, **11**(2):106-112.
3. Lobner-Olesen A, Skovgaard O, Marinus MG: **Dam methylation: coordinating cellular processes.** *Current Opinion in Microbiology* 2005, **8**(2):154-160.
4. Bickle TA, Kruger DH: **Biology of DNA Restriction.** *Microbiological Reviews* 1993, **57**(2):434-450.
5. Vasu K, Nagaraja V: **Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense.** *Microbiology and Molecular Biology Reviews* 2013, **77**(1):53-72.
6. Katayama T, Ozaki S, Keyamura K, Fujimitsu K: **Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC.** *Nature reviews Microbiology* 2010, **8**(3):163-170.
7. Modrich P, Lahue R: **Mismatch repair in replication fidelity, genetic recombination, and cancer biology.** *Annual review of biochemistry* 1996, **65**:101-133.
8. Braaten BA, Platko JV, Vanderwoude MW, Simons BH, Degraaf FK, Calvo JM, Low DA: **Leucine-Responsive Regulatory Protein Controls the Expression of Both the Pap and Fan Pili Operons in Escherichia-Coli.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(10):4250-4254.
9. Braun RE, Wright A: **DNA Methylation Differentially Enhances the Expression of One of the 2 Escherichia-Coli Dnaa Promoters In vivo and In vitro.** *Molecular & General Genetics* 1986, **202**(2):246-250.
10. Brunet YR, Bernard CS, Gavioli M, Lloubes R, Cascales E: **An epigenetic switch involving overlapping fur and DNA methylation optimizes expression of a type VI secretion gene cluster.** *Plos Genetics* 2011, **7**(7):e1002205-e1002205.
11. Camacho EM, Casadesus J: **Regulation of traJ transcription in Salmonella virulence plasmid by strand-specific DNA adenine hemimethylation.** *Molecular Microbiology* 2005, **57**(6):1700-1718.
12. Srikhanta YN, Fox KL, Jennings MP: **The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes.** *Nature Reviews Microbiology* 2010, **8**(3):196-206.
13. Hernday AD, Braaten BA, Low DA: **The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch.** *Molecular cell* 2003, **12**(4):947-957.
14. Lim HN, van Oudenaarden A: **A multistep epigenetic switch enables the stable inheritance of DNA methylation states.** *Nature Genetics* 2007, **39**(2):269-275.
15. Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS: **Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription.** *Nature communications* 2012, **3**:886.

- 547 16. Bormann Chung CA, Boyd VL, McKernan KJ, Fu Y, Monighetti C, Peckham  
548 HE, Barker M: **Whole methylome analysis by ultra-deep sequencing using**  
549 **two-base encoding.** *Plos One* 2010, **5**(2):e9320.
- 550 17. Ehrlich M, Wilson GG, Kuo KC, Gehrke CW: **N4-Methylcytosine as a Minor**  
551 **Base in Bacterial-DNA.** *J Bacteriol* 1987, **169**(3):939-943.
- 552 18. Wang MX, Church GM: **A Whole Genome Approach to Invivo DNA-Protein**  
553 **Interactions in Escherichia-Coli.** *Nature* 1992, **360**(6404):606-610.
- 554 19. Ringquist S, Smith CL: **The Escherichia-Coli Chromosome Contains Specific,**  
555 **Unmethylated Dam and Dcm Sites.** *Proceedings of the National Academy of*  
556 *Sciences of the United States of America* 1992, **89**(10):4539-4543.
- 557 20. Hale WB, Vanderwoude MW, Low DA: **Analysis of Nonmethylated Gate Sites**  
558 **in the Escherichia-Coli Chromosome and Identification of Sites That Are**  
559 **Differentially Methylated in Response to Environmental Stimuli.** *Journal of*  
560 *Bacteriology* 1994, **176**(11):3438-3441.
- 561 21. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach  
562 J, Turner SW: **Direct detection of DNA methylation during single-molecule,**  
563 **real-time sequencing.** *Nature Methods* 2010, **7**(6):461-U472.
- 564 22. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z,  
565 Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA,  
566 Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V,  
567 Waldor MK, Schadt EE: **Genome-wide mapping of methylated adenine**  
568 **residues in pathogenic Escherichia coli using single-molecule real-time**  
569 **sequencing.** *Nature Biotechnology* 2012, **30**(12):1232-1239.
- 570 23. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M,  
571 Fomenkov A, Roberts RJ, Korlach J: **Characterization of DNA**  
572 **methyltransferase specificities using single-molecule, real-time DNA**  
573 **sequencing.** *Nucleic Acids Res* 2012, **40**(4).
- 574 24. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov  
575 A, Turner SW, Korlach J, Roberts RJ: **The methylomes of six bacteria.** *Nucleic*  
576 *Acids Res* 2012, **40**(22):11450-11462.
- 577 25. Boukhalfa H, Icopini GA, Reilly SD, Neu MP: **Plutonium(IV) reduction by the**  
578 **metal-reducing bacteria Geobacter metallireducens GS15 and Shewanella**  
579 **oneidensis MR1.** *Applied and Environmental Microbiology* 2007, **73**(18):5897-  
580 5903.
- 581 26. Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner  
582 TS, Nealson KH, Osterman AL, Pinchuk G, Reed JL, Rodionov DA, Rodrigues  
583 JLM, Saffarini DA, Serres MH, Spormann AM, Zhulin IB, Tiedje JM: **Towards**  
584 **environmental systems biology of Shewanella.** *Nature reviews Microbiology*  
585 2008, **6**(8):592-603.
- 586 27. Myers CR, Nealson KH: **Bacterial manganese reduction and growth with**  
587 **manganese oxide as the sole electron acceptor.** *Science (New York, NY)* 1988,  
588 **240**(4857):1319-1321.
- 589 28. Caccavo F, Blakemore RP, Lovley DR: **A Hydrogen-Oxidizing, Fe(III)-**  
590 **Reducing Microorganism from the Great Bay Estuary, New-Hampshire.**  
591 *Applied and Environmental Microbiology* 1992, **58**(10):3211-3216.

- 592 29. Hau HH, Gralnick JA: **Ecology and biotechnology of the genus *Shewanella*.**  
593 *Annual Review of Microbiology* 2007, **61**:237-258.
- 594 30. Kolker E, Picone AF, Galperin MY, Romine MF, Higdon R, Makarova KS,  
595 Kolker N, Anderson GA, Qiu X, Auberry KJ, Babnigg G, Beliaev AS, Edlefsen P,  
596 Elias DA, Gorby YA, Holzman T, Klappenbach JA, Konstantinidis KT, Land  
597 ML, Lipton MS, McCue L-A, Monroe M, Pasa-Tolic L, Pinchuk G, Purvine S,  
598 Serres MH, Tsapin S, Zakrajsek BA, Zhu W, Zhou J, Larimer FW, Lawrence CE,  
599 Riley M, Collart FR, Yates JR, Smith RD, Giometti CS, Nealson KH, Fredrickson  
600 JK, Tiedje JM: **Global profiling of *Shewanella oneidensis* MR-1: expression of**  
601 **hypothetical genes and improved functional annotations.** *Proceedings of the*  
602 *National Academy of Sciences of the United States of America* 2005, **102**(6):2099-  
603 2104.
- 604 31. Price MN, Deutschbauer AM, Skerker JM, Wetmore KM, Ruths T, Mar JS, Kuehl  
605 JV, Shao WJ, Arkin AP: **Indirect and suboptimal control of gene expression is**  
606 **widespread in bacteria.** *Molecular Systems Biology* 2013, **9**.
- 607 32. Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I,  
608 Rodionov DA: **RegPrecise: a database of curated genomic inferences of**  
609 **transcriptional regulatory interactions in prokaryotes.** *Nucleic Acids Res*  
610 2010, **38**(Database issue):D111-118.
- 611 33. Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, Xu Z, Nguyen  
612 M, Tamse R, Davis RW, Arkin AP: **Evidence-based annotation of gene**  
613 **function in *Shewanella oneidensis* MR-1 using genome-wide fitness profiling**  
614 **across 121 conditions.** *Plos Genetics* 2011, **7**(11):e1002385-e1002385.
- 615 34. Heidelberg JF, Paulsen IT, Nelson KE, Gaidos EJ, Nelson WC, Read TD, Eisen  
616 JA, Seshadri R, Ward N, Methe B, Clayton RA, Meyer T, Tsapin A, Scott J,  
617 Beanan M, Brinkac L, Daugherty S, DeBoy RT, Dodson RJ, Durkin AS, Haft  
618 DH, Kolonay JF, Madupu R, Peterson JD, Umayam LA, White O, Wolf AM,  
619 Vamathevan J, Weidman J, Impraim M, Lee K, Berry K, Lee C, Mueller J,  
620 Khouri H, Gill J, Utterback TR, McDonald LA, Feldblyum TV, Smith HO,  
621 Venter JC, Nealson KH, Fraser CM: **Genome sequence of the dissimilatory**  
622 **metal ion-reducing bacterium *Shewanella oneidensis*.** *Nature Biotechnology*  
623 2002, **20**(11):1118-1123.
- 624 35. Roberts RJ, Vincze T, Posfai J, Macelis D: **REBASE-a database for DNA**  
625 **restriction and modification: enzymes, genes and genomes.** *Nucleic Acids Res*  
626 2010, **38**:D234-D236.
- 627 36. Gommers-Ampt JH, Borst P: **Hypermodified bases in DNA.** *Faseb J* 1995,  
628 **9**(11):1034-1042.
- 629 37. Iyer LM, Zhang D, Maxwell Burroughs A, Aravind L: **Computational**  
630 **identification of novel biochemical systems involved in oxidation,**  
631 **glycosylation and other complex modifications of bases in DNA.** *Nucleic Acids*  
632 *Res* 2013.
- 633 38. Dryden DTF, Murray NE, Rao DN: **Nucleoside triphosphate-dependent**  
634 **restriction enzymes.** *Nucleic Acids Res* 2001, **29**(18):3728-3741.
- 635 39. Lluch-Senar M, Luong K, Llorens-Rico V, Delgado J, Fang G, Spittle K, Clark  
636 TA, Schadt E, Turner SW, Korlach J, Serrano L: **Comprehensive methylome**

637 characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at  
638 single-base resolution. *Plos Genetics* 2013, 9(1):e1003191.

639 40. Tavazoie S, Church GM: **Quantitative whole-genome analysis of DNA-protein**  
640 **interactions by in vivo methylase protection in E-coli.** *Nature Biotechnology*  
641 1998, 16(6):566-571.

642 41. Low DA, Weyand NJ, Mahan MJ: **Roles of DNA adenine methylation in**  
643 **regulating bacterial gene expression and virulence.** *Infection and Immunity*  
644 2001, 69(12):7197-7204.

645 42. Oshima T, Wada C, Kawagoe Y, Ara T, Maeda M, Masuda Y, Hiraga S, Mori H:  
646 **Genome-wide analysis of deoxyadenosine methyltransferase-mediated**  
647 **control of gene expression in Escherichia coli.** *Molecular Microbiology* 2002,  
648 45(3):673-695.

649 43. Lobner-Olesen A, Marinus MG, Hansen FG: **Role of SeqA and Dam in**  
650 **Escherichia coli gene expression: A global/microarray analysis.** *Proceedings*  
651 *of the National Academy of Sciences of the United States of America* 2003,  
652 100(8):4672-4677.

653 44. Robbins-Manke JL, Zdraveski ZZ, Marinus M, Essigmann JM: **Analysis of**  
654 **global gene expression and double-strand-break formation in DNA adenine**  
655 **methyltransferase-and mismatch repair-deficient Escherichia coli.** *Journal of*  
656 *Bacteriology* 2005, 187(20):7027-7037.

657 45. Casadesus J, Low DA: **Programmed heterogeneity: epigenetic mechanisms in**  
658 **bacteria.** *J Biol Chem* 2013, 288(20):13929-13935.

659 46. Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N: **E. coli SeqA protein**  
660 **binds oriC in two different methyl-modulated reactions appropriate to its**  
661 **roles in DNA replication initiation and origin sequestration.** *Cell* 1995,  
662 82(6):927-936.

663 47. Boye E, Lobner-Olesen A, Skarstad K: **Limiting DNA replication to once and**  
664 **only once.** *Embo Reports* 2000, 1(6):479-483.

665 48. Oka A, Sugimoto K, Takanami M, Hirota Y: **Replication Origin of the**  
666 **Escherichia-Coli K-12 Chromosome - the Size and Structure of the Minimum**  
667 **DNA Segment Carrying the Information for Autonomous Replication.**  
668 *Molecular and General Genetics* 1980, 178(1):9-20.

669 49. Ogden GB, Pratt MJ, Schaechter M: **The Replicative Origin of the Escherichia-**  
670 **Coli Chromosome Binds to Cell-Membranes Only When Hemimethylated.**  
671 *Cell* 1988, 54(1):127-135.

672 50. Vonfreiesleben U, Rasmussen KV, Schaechter M: **Seqa Limits Dnaa Activity in**  
673 **Replication from Oric in Escherichia-Coli.** *Molecular Microbiology* 1994,  
674 14(4):763-772.

675 51. Lu M, Campbell JL, Boye E, Kleckner N: **Seqa - a Negative Modulator of**  
676 **Replication Initiation in Escherichia-Coli.** *Cell* 1994, 77(3):413-426.

677 52. Boye E, Lobner-Olesen A: **The role of dam methyltransferase in the control of**  
678 **DNA replication in E. coli.** *Cell* 1990, 62:981-989.

679 53. van Opijnen T, Bodi KL, Camilli A: **Tn-seq: high-throughput parallel**  
680 **sequencing for fitness and genetic interaction studies in microorganisms.**  
681 *Nature Methods* 2009, 6(10):767-772.

- 682 54. Julio SM, Heithoff DM, Provenzano D, Klose KE, Sinsheimer RL, Low DA,  
683 Mahan MJ: **DNA adenine methylase is essential for viability and plays a role**  
684 **in the pathogenesis of *Yersinia pseudotuberculosis* and *Vibrio cholerae*.**  
685 *Infection and Immunity* 2001, **69**(12):7610-7615.
- 686 55. Demarre GI, Chattoraj DK: **DNA adenine methylation is required to replicate**  
687 **both *Vibrio cholerae* chromosomes once per cell cycle.** *Plos Genetics* 2010,  
688 **6**(5):e1000939-e1000939.
- 689 56. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A,  
690 Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K,  
691 Ivanova NN, Kyrpides NC: **IMG: the Integrated Microbial Genomes database**  
692 **and comparative analysis system.** *Nucleic Acids Res* 2012, **40**(Database  
693 issue):D115-122.
- 694 57. Moran NA: **Microbial Minimalism: Genome Reduction in Bacterial**  
695 **Pathogens.** *Cell* 2002, **108**(5):583-586.
- 696 58. Bian F, Qin Q-L, Xie B-B, Shu Y-L, Zhang X-Y, Yu Y, Chen B, Chen X-L, Zhou  
697 B-C, Zhang Y-Z: **Complete genome sequence of seawater bacterium**  
698 ***Glaciecola nitratreducens* FR1064(T).** *Journal of Bacteriology* 2011,  
699 **193**(24):7006-7007.
- 700 59. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW: **A flexible and efficient**  
701 **template format for circular consensus sequencing and SNP detection.**  
702 *Nucleic Acids Res* 2010, **38**(15):e159.
- 703 60. Chaisson MJ, Tesler G: **Mapping single molecule sequencing reads using basic**  
704 **local alignment with successive refinement (BLASR): application and theory.**  
705 *Bmc Bioinformatics* 2012, **13**:238.
- 706 61. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobisch S,  
707 Lehrach H, Soldatov A: **Transcriptome analysis by strand-specific sequencing**  
708 **of complementary DNA.** *Nucleic Acids Res* 2009, **37**(18).
- 709 62. Anders S, Huber W: **Differential expression analysis for sequence count data.**  
710 *Genome Biology* 2010, **11**(10):R106.
- 711 63. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for**  
712 **differential expression analysis of digital gene expression data.** *Bioinformatics*  
713 2010, **26**(1):139-140.
- 714 64. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J:  
715 **RankProd: a bioconductor package for detecting differentially expressed**  
716 **genes in meta-analysis.** *Bioinformatics* 2006, **22**(22):2825-2827.
- 717 65. Gao F, Luo H, Zhang CT: **DoriC 5.0: an updated database of oriC regions in**  
718 **both bacterial and archaeal genomes.** *Nucleic Acids Res* 2013, **41**(Database  
719 issue):D90-93.
- 720 66. Gao F, Zhang CT: **DoriC: a database of oriC regions in bacterial genomes.**  
721 *Bioinformatics* 2007, **23**(14):1866-1867.
- 722 67. Gao F, Zhang CT: **Ori-Finder: a web-based system for finding oriCs in**  
723 **unannotated bacterial genomes.** *Bmc Bioinformatics* 2008, **9**:79.
- 724 68. Katoh K, Toh H: **Recent developments in the MAFFT multiple sequence**  
725 **alignment program.** *Briefings in bioinformatics* 2008, **9**(4):286-298.

726 69. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic**  
727 **analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006,  
728 **22(21):2688-2690.**

729 70. Gao B, Mohan R, Gupta RS: **Phylogenomics and protein signatures**  
730 **elucidating the evolutionary relationships among the Gammaproteobacteria.**  
731 *International journal of systematic and evolutionary microbiology* 2009, **59**(Pt  
732 **2):234-247.**

733 71. Williams KP, Gillespie JJ, Sobral BW, Nordberg EK, Snyder EE, Shallom JM,  
734 Dickerman AW: **Phylogeny of gammaproteobacteria.** *Journal of Bacteriology*  
735 2010, **192**(9):2305-2314.

736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771

Table 1. Sequence motifs containing N6mA modifications (underlined bases) in *S. oneidensis* MR-1 grown aerobically on minimal media. The total number of motifs and their location in either coding (CDS) or intergenic regions was determined. The methylation state of motifs with <25X coverage could not be confidently determined and were excluded from these counts. 'Methylated' motifs were modified on both strands, while 'Hemi' motifs were modified on only one strand. Motifs with disagreements among replicates were in 'Conflict'. The percentage of total motifs methylated is indicated in ()'s. Methylated bases occurring simultaneously in two motifs were counted towards each motif, e.g. the 225 instances of the methylated sequence ATCGATC were included in the tallies of both GATC and ATCGAT motifs.

| Motif                                             | Total  | CDS    | Intergenic | Methylation state |      |     |          |  |
|---------------------------------------------------|--------|--------|------------|-------------------|------|-----|----------|--|
|                                                   |        |        |            | Methylated        | Hemi | Not | Conflict |  |
| GATC                                              | 16,376 | 14,328 | 2,048      | 16,281 (99.4%)    | 0    | 2   | 93       |  |
| ATCGAT                                            | 2,342  | 2,052  | 290        | 2,292 (97.9%)     | 3    | 1   | 46       |  |
| GCAN <sub>4</sub> GTC / GACN <sub>4</sub> TGC     | 1,803  | 1,543  | 260        | 1,793 (99.4%)     | 1    | 0   | 9        |  |
| TACN <sub>6</sub> GTNGT / ACNACN <sub>6</sub> GTA | 306    | 246    | 60         | 301 (98.4%)       | 0    | 0   | 5        |  |
| TGAYN <sub>6</sub> TGAC / GTCAN <sub>6</sub> RTCA | 291    | 243    | 48         | 282 (96.9%)       | 1    | 5   | 3        |  |

773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798

799  
800

Table 2. Assigning target motifs to putative methyltransferases using SMRT sequencing of gene knockout mutants. Underlined bases were methylated. Target motifs were predicted for three putative methyltransferase in REBASE. Loss of a methylated motif in mutants confirmed the target of the knocked-out gene. Motifs for three predicted methyltransferases were not detected (ND). No mutant was available for SO\_0289, whereas no methylated motifs were lost for SO\_0690 and SO\_3004.

| Locus ID | R-M System | Predicted motif | Confirmed motif                                                     |
|----------|------------|-----------------|---------------------------------------------------------------------|
| SO_0383  | Type I     |                 | TG <u>A</u> YN <sub>6</sub> TGAC / GTC <u>A</u> N <sub>6</sub> RTCA |
| SO_1457  | Type I*    |                 | G <u>C</u> AN <sub>4</sub> GTC / G <u>A</u> CN <sub>4</sub> TGC     |
| SO_4265  | Type I     |                 | T <u>A</u> CN <sub>6</sub> GTNGT / ACN <u>A</u> CN <sub>6</sub> GTA |
| SO_0289  | Type II    | GATC            | ND                                                                  |
| SO_0690  | Type II    |                 | ND                                                                  |
| SO_3004  | Type II    | GATC            | ND                                                                  |
| SO_A0004 | Type II    | ATCGAT          | ATCG <u>A</u> T                                                     |

\* Type II according to REBASE, but annotated as Type I elsewhere (e.g. Genbank, COG)

801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820

821  
822

Table 3. Transcription factor binding sites containing Type II motifs. The number of genes and operons regulated by these transcription factors, as well as the number of genes differentially expressed when transferred to either Rich media or Anaerobic minimal media, is indicated. The number of up-regulated genes is indicated in ()'s next to the total number of differentially expressed genes. Bindings sites with different methylation states under different growth conditions are also displayed.

| Regulator | Total Sites | Sites with motif | Operons regulated | Genes regulated | Differentially expressed genes |           | Differentially methylated sites |           |
|-----------|-------------|------------------|-------------------|-----------------|--------------------------------|-----------|---------------------------------|-----------|
|           |             |                  |                   |                 | Rich media                     | Anaerobic | Rich media                      | Anaerobic |
| ArgR      | 34          | 1                | 22                | 39              | 30 (4)                         | 13 (11)   | 0                               | 0         |
| Crp       | 174         | 64               | 150               | 286             | 69 (48)                        | 12 (6)    | 0                               | 0         |
| Fnr       | 30          | 21               | 26                | 73              | 34 (28)                        | 8 (6)     | 0                               | 0         |
| Fur       | 32          | 1                | 32                | 58              | 12 (7)                         | 7 (2)     | 1                               | 0         |
| ModE      | 2           | 2                | 2                 | 4               | 0                              | 0         | 0                               | 0         |
| SO1578    | 2           | 2                | 2                 | 3               | 0                              | 0         | 0                               | 0         |
| SO3385    | 1           | 1                | 1                 | 3               | 0                              | 0         | 0                               | 0         |
| SO3393    | 6           | 2                | 2                 | 2               | 1 (1)                          | 0         | 0                               | 0         |

823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848

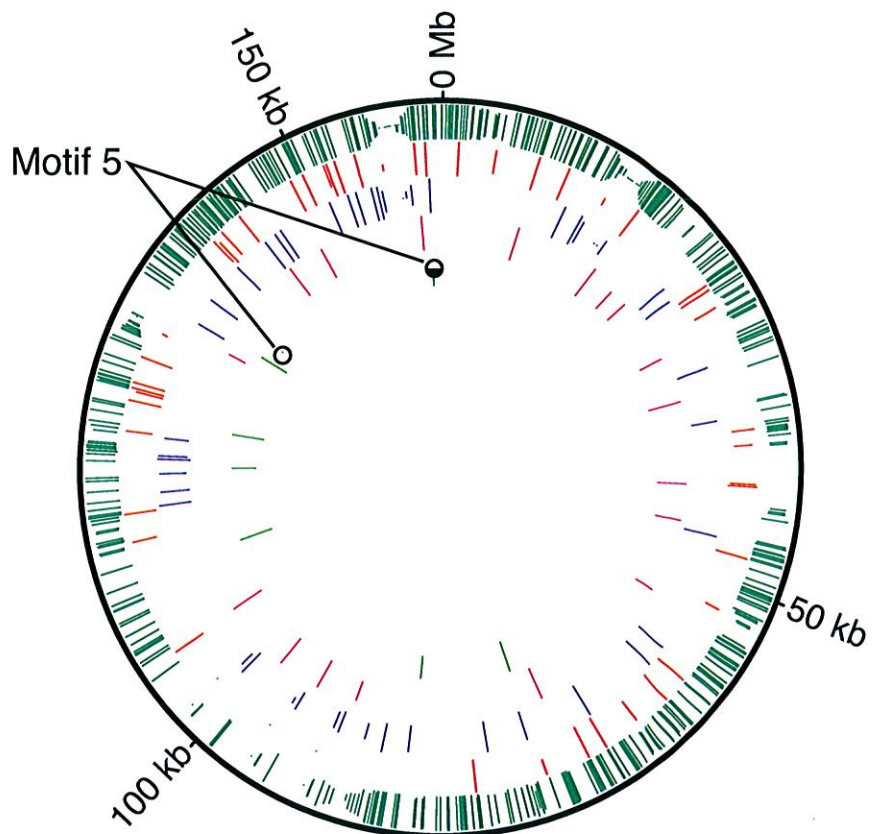
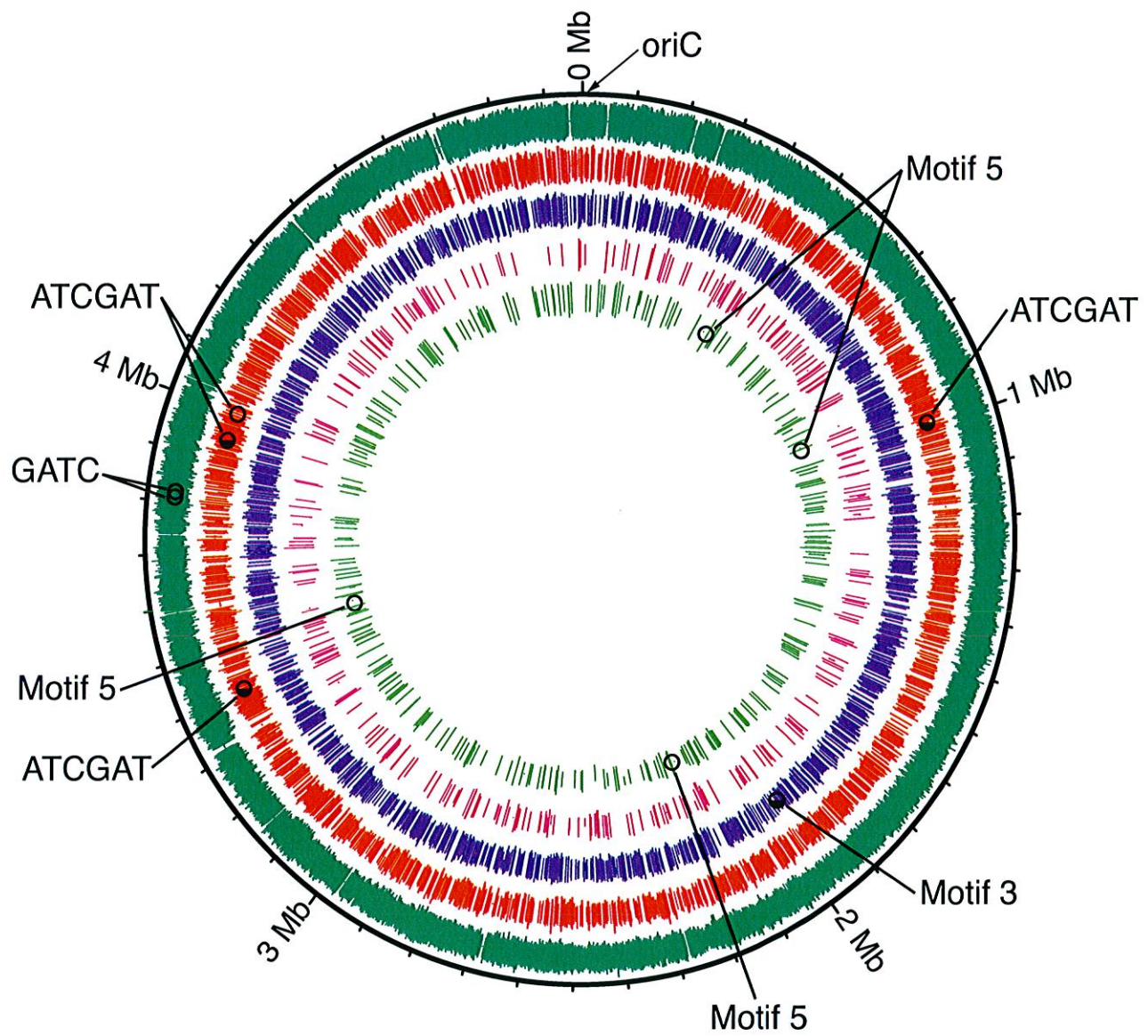
849  
850  
851

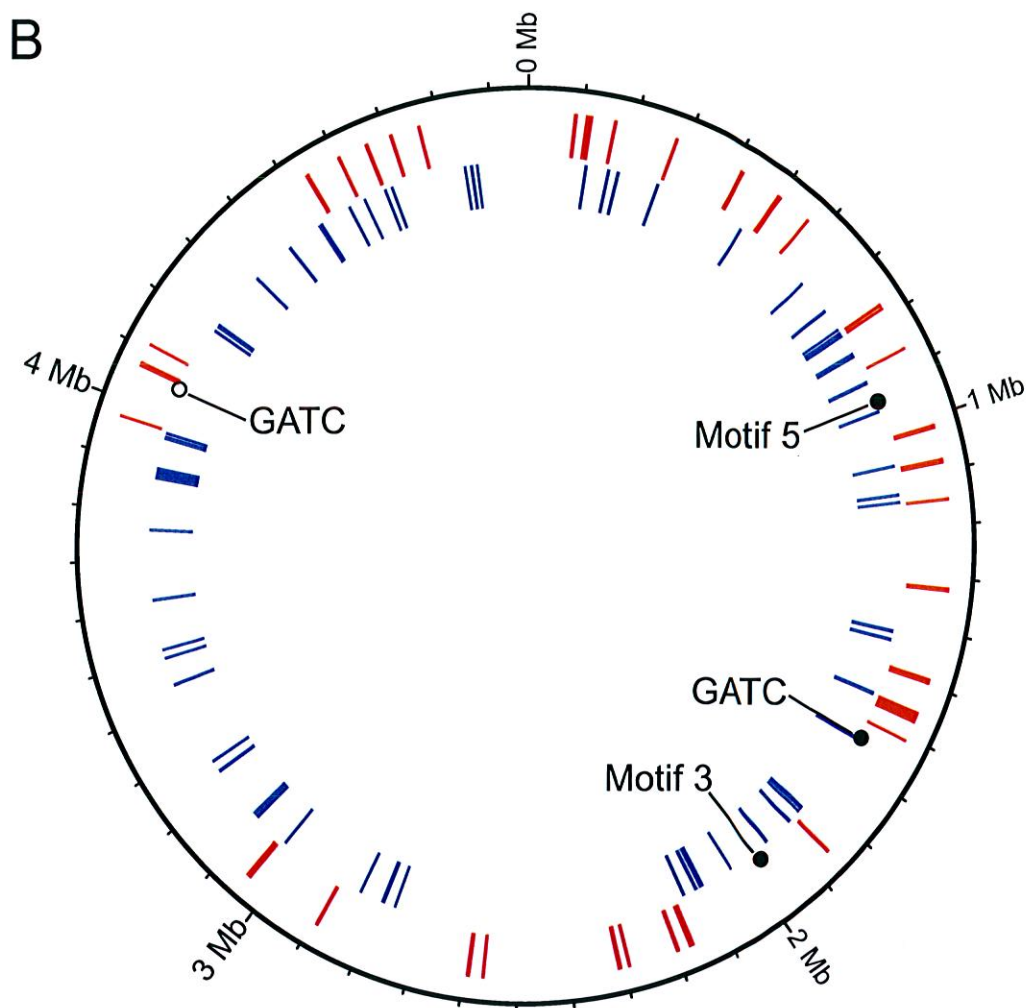
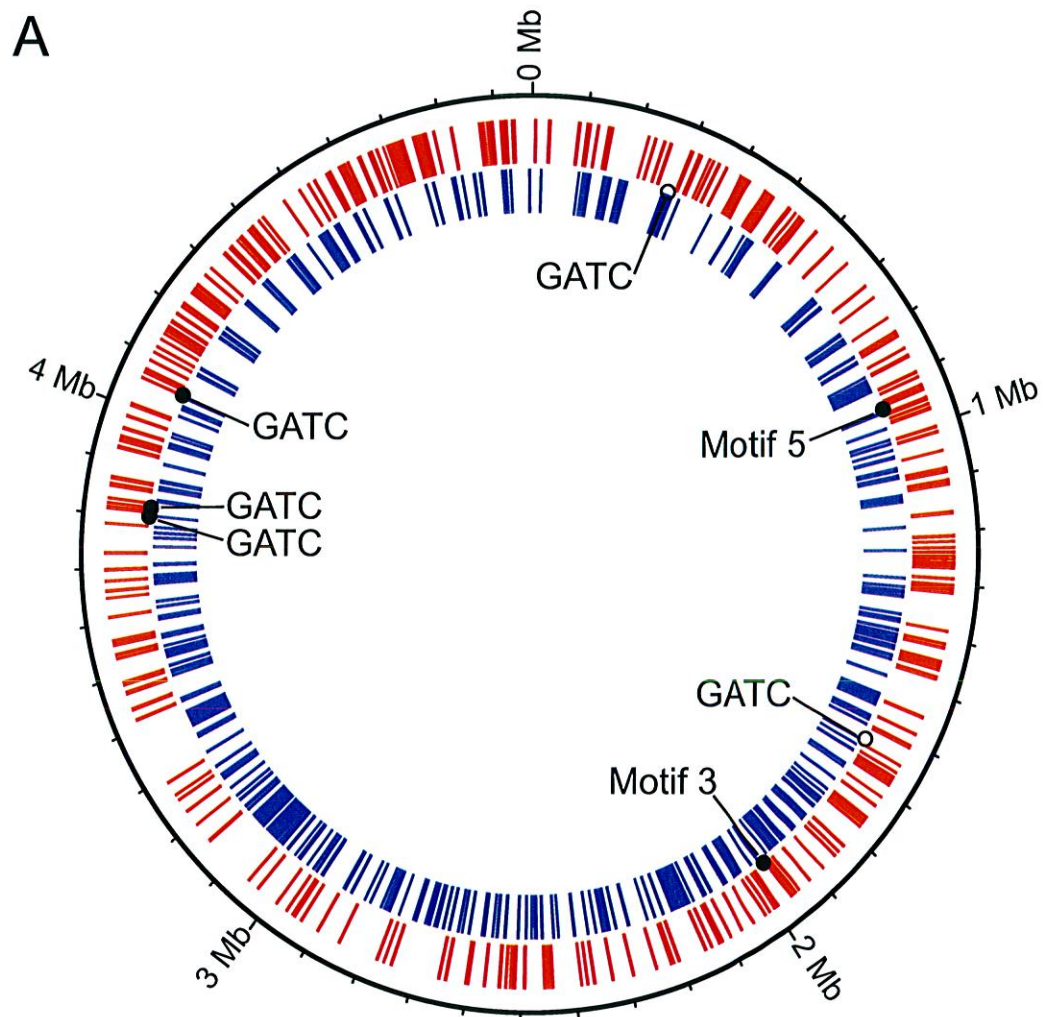
Table 4. Genomic context of motifs with differences in methylation state among conditions. Left and right flanking genes, as well as their orientation, are provided for motifs found within intergenic regions. Motifs were down stream of a left flanking gene if the arrow points to the left, and down stream of right flanking genes if the arrow points to the right. "P" indicates motif located in the plasmid, while \* indicated motif in a Fur transcription factor binding site.

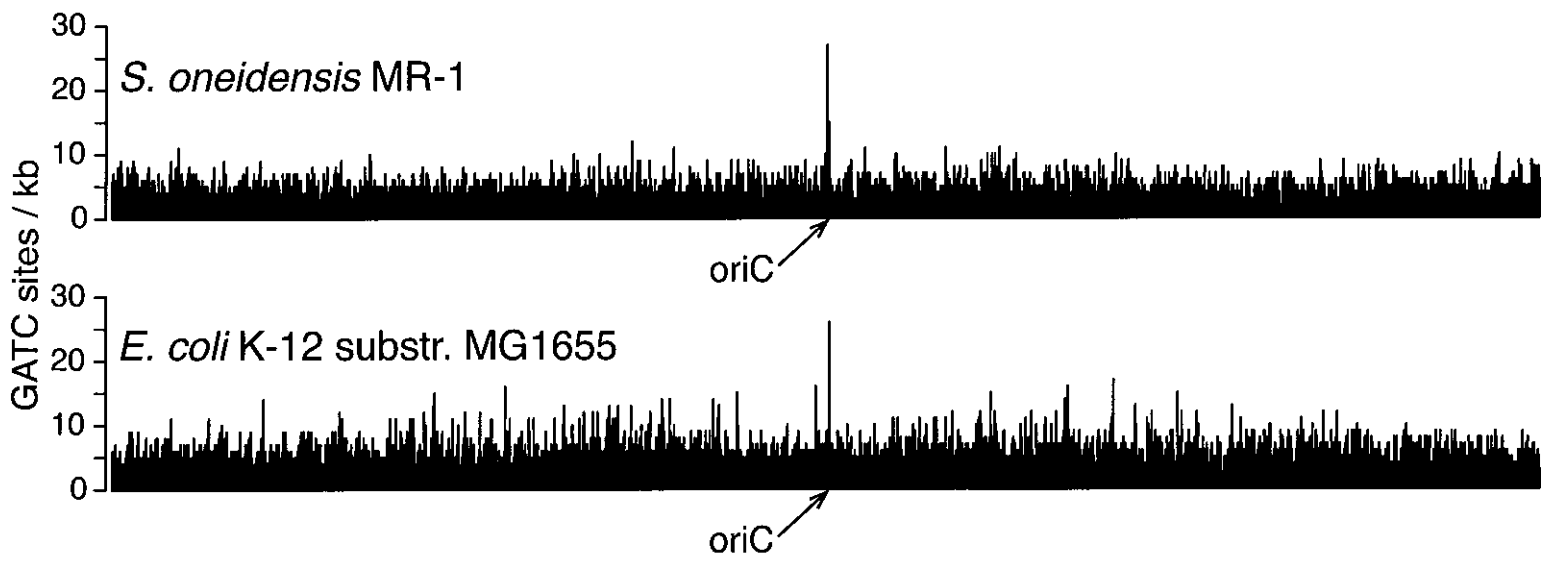
| Position (+/-)   | Motif                                              | Methylation state |            |                     | Genomic context    |             |                     |
|------------------|----------------------------------------------------|-------------------|------------|---------------------|--------------------|-------------|---------------------|
|                  |                                                    | Min Media Aerobic | Rich media | Min Media Anaerobic | Left flanking gene | Within gene | Right flanking gene |
| 280309/280310    | GATC                                               | ●                 | ○          | ●                   | ← SO_0274          | -           | SO_0275 →           |
| 1642124/1642125  | GATC                                               | -                 | ○          | ●                   | ← SO_1563          | -           | SO_1565 →           |
| 3823765/3823766  | GATC                                               | ○                 | ●          | ○                   | ← SO_3669          | -           | SO_3670 →           |
| 3823792/3823793  | GATC                                               | ○                 | ●          | ○                   | ← SO_3669          | -           | SO_3670 →           |
| 4061174/4061175* | GATC                                               | -                 | ●          | ○                   | ← SO_3914          | -           | SO_3915 →           |
| 1965318/1965324  | GCA <sub>N</sub> 4GTC /<br>GAC <sub>N</sub> 4TGC   | ●                 | ●          | ●                   | → SO_1871          | -           | SO_1872 ←           |
| 938801/938809    | TGA <sub>N</sub> 6TGAC /<br>GTC <sub>N</sub> 6RTCA | ○                 | ●          | ●                   | -                  | SO_0912     | -                   |
| 160490/160498 P  | TGA <sub>N</sub> 6TGAC /<br>GTC <sub>N</sub> 6RTCA | ●                 | -          | ●                   | -                  | SO_A0172    | -                   |

● = methylated      ○ = non-methylated      ● = hemi-methylated

852  
853  
854







- dam + seqA + enrichment
- dam only
- none

