Supplemental Text.

Expanded Materials and Methods.

Genomic DNA preparation. Cells were washed once with TE buffer (10 mM Tris•HCl, pH 8.0; 1 mM EDTA, pH 8.0) and resuspended in 180 µL enzymatic lysis buffer (20 mM Tris•HCl, pH 8.0; 2 mM sodium EDTA; 1.2% Triton X-100) amended with 20 mg/mL lysozyme, 10 µL of a 2.5 kU/mL mutanolysin stock, and 15 µL of a 10 mg/mL pre-boiled RNase A stock. Samples were incubated at 37°C for 1-2 hours prior to proteinase K treatment and column purification of gDNA using the Qiagen DNeasy Blood and Tissue kit per the manufacturer's instructions. *E. coli* gDNA was isolated using the same method except mutanolysin was not included.

Pacific Biosciences SMRT sequencing. SMRT sequencing of native and WGA *E. faecalis* OG1RF gDNA was performed at the University of California San Diego BIOGEM core facility. 800 bp insert libraries were built using the Sequencing kit 2.0 as described in the SMRTbell Template Preparation workflow (1). Libraries were sequenced using 2 x 45 minute movies and 3 SMRT cells for OG1RF native gDNA and 4 SMRT cells for OG1RF WGA DNA. SMRT sequencing reads were assembled to the *E. faecalis* OG1RF reference sequence (Genbank accession NC_017316). Mean read lengths obtained were 3227 and 3760 bases for OG1RF native and WGA DNA, respectively.

SMRT sequencing of *E. coli* BL21DE3 gDNA after overexpression of predicted MTases was performed at the University of Michigan sequencing core facility using P6-C4 chemistry. The gDNA of *E. coli* BL21DE3 with induced empty vector pET28a was sequenced as negative control. Two SMRT cells were sequenced for each sample. SMRT sequencing reads were assembled to the *E. coli* BL21DE3 reference sequence (Genbank accession CP001509.3).

Mean read lengths obtained were 9315, 5745 and 8557 bases for *E. coli* BL21DE3 containing induced pET28a, pWH21 and pWH51, respectively.

The interpulse duration (IPD) is the time elapsed between incorporation of adjacent nucleotides by DNA polymerase, and the IPD ratio refers to the ratio of IPD values between native and control templates for a given nucleotide position. The significance of the IPD ratio is evaluated using Welch's t-test, with the resulting p-value further transformed into a quality value (QV; QV = -10log p-value; details can be found at https://github.com/PacificBiosciences/kineticsTools). The IPD data generated from the control (OG1RF WGA or BL21DE3 with induced pET28a) library were analyzed using the RS_Modification_and_Motif_Analysis.1 protocol in the SMRT Portal (v1.3.3 or v2.3.0, respectively) using default analysis parameters. The IPD data generated from the native gDNA library were then analyzed using the same protocol with the control result as post-processing. This calculated the IPD ratio at each genomic position by dividing the IPD from native DNA with that from control DNA. Modified sequence motifs were also predicted by the analysis. Additional analyses of motif enrichment were performed with MEME (2).

Illumina MiSeq whole genome bisulfite sequencing. Bisulfite-converted Illumina sequencing libraries were constructed using the Illumina TruSeq Sample Prep (LT) Kit and the Qiagen EpiTect Bisulfite kit. *E. coli* strains BW25113 and JW1944 (*dcm*-) (3, 4) were used as positive and negative controls for bisulfite conversion due to the well characterized Dcm methylation system. The *dcm*- mutation in JW1944 was confirmed by PCR (data not shown). OG1RF WGA control DNA was also used as negative control. Briefly, DNA was fragmented using a Bioruptor Sonicator UCD-200, and 200-500 bp fragments were gel-extracted and end-repaired. After adapter ligation, bisulfite conversion was performed using the Qiagen EpiTect Bisulfite kit per the manufacturer's instructions, followed by an eight cycle PCR amplification using Taq

polymerase (NEB) and a primer cocktail provided by the TruSeq Kit. Illumina sequencing was performed on an Illumina MiSeq with 2x150 paired end sequencing.

Illumina sequence reads were mapped to the OG1RF reference sequence or the *E. coli* K12 MG1655 reference sequence (GenBank accession number: NC_000913) using Bismark (5) with paired-end read mapping based on Bowtie2 (6). In read mappings, every C position could be represented either as a C (protected from bisulfite conversion) or a thymidine (T; bisulfite-converted cytosine). We observed that bisulfite conversion failed for a subset of input DNA (discussed further below). To correct for this, we determined the bisulfite conversion rate for individual reads by dividing the number of C-to-T conversions in each mapped read by the total number of Cs in the corresponding reference sequence. For determination of methylation ratios, only reads with ≥80% bisulfite conversion rate were used. A sequencing depth threshold of ≥7 was further applied to reduce bias generated by low coverage.

The methylation ratio for each C position in the reference genome was calculated by dividing the number of Cs mapped by the total coverage at that position. The significance of the methylation ratio was calculated using empirical modeling, with OG1RF WGA or *dcm*- DNA as the background (negative control) observation. Briefly, our null hypothesis is that no C methylation exists in native/*dcm*+ DNA, meaning that the distributions of methylation ratios in native/*dcm*+ and control DNA are the same. To assess whether a given methylation ratio in native/*dcm*+ DNA significantly differs from background, we calculated the occurrence of C positions from negative control with larger than or equal to the given methylation ratio in native DNA, which was further divided by the total number of C positions from the negative control. Zero frequency was represented by one over the total number of C positions from the negative control. The proportion of C positions in the negative control with methylation ratios as extreme as those in native DNA represents an empirical p-value. The lower the p-value, the more

significantly the methylation ratio is different from background. The quality value was -10log transformed p-value. Quality values larger than 40 or empirical P-values smaller than e^{-4} were considered as significant methylated positions. The detection rate of 5'-C^{5m}CWGG-3' in *E. coli* strain indicated how well bisulfite sequencing detected m5C.

Expanded analysis of SMRT and bisulfite sequencing data.

Expanded analysis of *E. faecalis* SMRT sequencing data. Table 3 in the main text shows the modified sequence motifs predicted by the SMRT analysis pipeline for *E. faecalis* OG1RF. To explore the C modification motifs further, we extracted the nucleotide positions with the top 1000 QVs (most significant IPD ratios) along with 20 bases of upstream and downstream sequence. The sequences were analyzed for putative modification motifs with MEME (30). For 560 nucleotide positions with high QVs, a consensus GCAGC motif was identified two positions downstream, which suggests that the high QV positions are actually secondary signals resulting from methylation at GCAGC motifs (Fig. S2). For 45 nucleotide positions with high QVs, a consensus 5'-GCWGC-3' motif was identified five positions downstream (Fig. S2). For 78 positions with high QV, a GCTGC motif was identified within surrounding sequence, although the position relative to the high QV base varied (Fig. S2). No motifs were detected by MEME for the remaining 317 high QV positions. Broadening this analysis, for the 3610 positions with QV >40, 1533 positions have a 5'-GCWGC-3' motif nearby (within 20 bases up- or downstream), out of which 461 positions are surrounded by >1 5'-GCWGC-3' motifs (Table S2B). To further evaluate the secondary peaks associated with 5'-GCWGC-3' motifs, sequences surrounding 8326 5'-GCWGC-3' motifs were extracted and mean QVs for each position were calculated (Fig. S2). The underlined positions within motif 5'-$\underline{N}(N_2)\underline{N}(N_2)\underline{N}\underline{N}G\underline{C}$WGC-3' have significantly higher QVs than the remaining positions (p-value = $6.31e^{-12}$). Based on the low detection rate and

104 secondary kinetic signals surrounding 5'-GCWGC-3' motifs, we predicted that these motifs were
105 modified with m5C.
106
107 **Expanded analysis of *E. faecalis* bisulfite sequencing data.** We used bisulfite sequencing to
108 explore the extent of m5C modification in the OG1RF genome, with OG1RF WGA DNA and *E.*
109 *coli dcm*+ and *dcm*- strains serving as controls. Illumina reads were mapped to reference
110 sequences, and visual inspection of the read mappings revealed that some reads contained no
111 bisulfite conversion at any C position (data not shown). This suggested that bisulfite conversion
112 had failed for some input DNA. Failure of bisulfite conversion for a subset of input DNA could
113 strongly bias the identification of methylated Cs.
114
115 To address this confounding factor, we evaluated the bisulfite conversion rate for individual
116 reads. The bisulfite conversion rate for each read was calculated by dividing the number of C-to-
117 T conversions in the read by the total number of Cs in the corresponding reference sequence.
118 By this calculation, a read for which bisulfite conversion failed would have a bisulfite conversion
119 rate of 0, and a read for which every C position was bisulfite converted would have a bisulfite
120 conversion rate of 1. Reads containing protected positions (m5C or m4C positions) would be
121 expected to have a bisulfite conversion rate between 0 and 1. Compared to the high conversion
122 rate for WGA control DNA and *E. coli* gDNA, the conversion rate per read varies for native
123 OG1RF gDNA (Fig. S4). Assuming the absence of methylated CpG islands or analogous
124 features, it is unlikely that reads with bisulfite conversion rates near 0 resulted from DNA that
125 was methylated at every C position in a 150 base region. We conclude that bisulfite conversion
126 failed for some input DNA.
127
128 To account for failed bisulfite conversion in a subset of input DNA, reads with bisulfite
129 conversion ratios <0.8 (<80% bisulfite conversion rate per read) were removed from analysis.

The number of mapped reads was reduced after application of this filter (Table S2C); however, the number of C positions covered was maintained (Table S2D), which suggests an unbiased distribution of sequencing reads with a low conversion rate. Figure 3 in the main text shows the distribution of methylation ratios obtained using these filtered assemblies.

Expanded analysis of *E. coli* SMRT sequencing data. The IPD ratio for BL21DE3 with induced pET28a was calculated using an *in silico* control, which predicts host modifications. For BL21DE3 with pET28a, we were able to detect 5'-GATC-3' with m6A modification successfully, modified by DNA adenine methyltransferase (*dam*) (see Table 3 in main text). The IPD ratio for BL21DE3 with induced pWH21 and pWH51 was calculated using the pET28a result as the control, which removed baseline modifications (i.e., Dam modification). One cytosine modification motif was detected for the strain expressing OG1RF_11823 (5'-CCGG-3', with underlined nucleotide modified). No modification motifs were detected for the strain expressing OG1RF_10790.

To further explore the modification type, we extracted the nucleotide positions with top 1000 highest QVs as described above for our *E. faecalis* SMRT analysis. The sequence contexts surrounding those nucleotide positions were analyzed by MEME. For 590 nucleotide positions, a consensus motif was found to be enriched in the center, indicating a strong primary signal within 5'-CCGG-3' at the underlined position (Fig. S2). The MEME analysis for the remaining 410 nucleotide positions did not detect a consensus sequence. Based on the center position enrichment from the MEME analysis, and the fact that this cytosine methylation could be detected at a relatively lower mean coverage depth (86X), we predicted this modification to be m4C.

For the strain expressing OG1RF_10790, SMRT analysis failed to detect modification motifs. However, the presence of m5C DNA modification cannot be excluded, due to insufficient coverage depth. We extracted the nucleotide positions with the top 1000 highest quality values and their surrounding sequence contexts and performed a MEME analysis. We were unable to identify a consensus sequence from these data.

One concern in using *E. coli* BL21DE3 as the MTase expression host is that the overexpressed MTases could modify 5'-GATC-3' sequences, which could not be distinguished from the host Dam modification. To eliminate this possibility, we compared the protein sequences of OG1RF_11823 and OG1RF_10790 to Dam using pairwise alignment. The proteins share only 12.2% and 13.0% amino acid sequence identity, respectively. Based on this and the REBASE predictions for these enzymes (Table 2 in the main text), we exclude the possibility that OG1RF_10790 and OG1RF_11823 modify 5'-GATC-3'.

References Cited.

1. **Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW.** 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Research **38:**e159.
2. **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.** 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research **37:**W202-208.
3. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology **2:**2006 0008.
4. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America **97:**6640-6645.
5. **Krueger F, Andrews SR.** 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27:**1571-1572.
6. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods **9:**357-359.

Table S1. Primers used.

Primer name	Sequence
pLT06 MCS for	TCGCATCCGATTGCAGTATAAA
pLT06 MCS rev	ACCTATCACCTCAAATGGTTCG
pAT28 For	CGTATGTTGTGTGGAATTGTG
pAT28 Rev	GGCGATTAAGTTGGGTAACG
clpX For	CGCACACTTTCTGTTGCTG
clpX Rev	CCATCAAATGCTCCACCAAC
16S *E. faecalis* For	CATGCAAGTCGAACGCTTCT
16S *E. faecalis* Rev	CCATATATCTACGCATTTCAC
Methyl Primers	
OG1RF_11844_For	GATATTTGTAGGTATATTTTTGTTT
OG1RF_11844_Rev	AAATATTACTTTCATTAATAAAACC
ECD_00002_For	TTTTGTATGGTATTAGTTTGTTGGG
ECD_00002_Rev	TACCATCAACACCATATAATCAACC
Generation of OG1RF ΔEfsI	
1F_BamHI	AATGAAGGATCCACCTAAAGCTATTCCTCCTGGA
1R_XbaI	AGTACATCTAGATCAATCTTAAAAGGTCGTGGCT
2F_PstI	TCATCTCTGCAGACTCAACAGATAAAGCATCCCC
2R_BamHI	TTATCTGGATCCGGAAGATTGGATGTAGAGATAACA
Confirmation of OG1RF ΔEfsI	
5F	ACATGCGTAAATCATTGATGTCA
5R	TATTCTTGGTGTGTCTATCGCC
3F	GCGTACAGCGTTATCTAAAACC
3R	CATTAAGCCGTACTGACCGTAT

4F	GAACTCCAAAATGGCAATGGAT
4R	TCTAAGCCGTCTTCGATTGAAA
Construction of pM.EfaRFI	
M1_F_EcoRI	ccgGAATTCCGATGTAGCTGAATACAAAGGC
M1_R_BamHI	cgcGGATCCCAACAGGTTCTATAAGCACTTT
Construction of pCom02	
RBS_R_NotI	AAGGAAAAAAGCGGCCGCTGGAAAATAGTGTCAAGAGAAGGA
ORF_F_NotI	AAGGAAAAAAGCGGCCGCTTAACCAAAAGGATTAAAATCTAAAT
Construction of pWH02	
addProRM Arm1 For PstI	aaaactgcagTTTGACTAATTTTTGCCCCTGC
addProRM Arm1 overlap Rev	AAAAAATAAGGACGGTTCCTTTATAGGAGCTTCATTTAATGAATAACGCTTAAAGGGAC
ef2239 R BamHI	CGCGGATCCGGTAAAACTAGGAGGGAAGCATATG
ef2239 F	GCTCCTATAAAGGAACCGTCCTTATTTTTT
Construction of pWH03	
ef2239 F	GCTCCTATAAAGGAACCGTCCTTATTTTTT
ef2238 R NotI	AAAAAATAAGGACGGTTCCTTTATAGGAGCGCGGCCGCGGTCCTTATTTTTTATTTCTGGCGTGG
ef2239 R XbaI	CTAGTCTAGAGGTAAAACTAGGAGGGAAGCATATG
ef2238 F PstI	aaaactgcagCTATAATAGTACTTGAGAAGGAGGC
In vitro expression	
OG1RF_11823_For_BamHI	CGCGGATCCATGAGGACTTTTATGGAAAGTATAA
OG1RF_11823_Rev_NotI	ATAAGAATGCGGCCGCTTATTTCCTCCCTACAACGATA
OG1RF_10790_For_NdeI	GGAATTCCATATGGTGGAATTTTTAGATTTATTTGCC
OG1RF_10790_Rev_BamHI	CGCGGATCCTCAATCCTTTAATTGTCTGGCAAT
pCF10 transconjugant verification	
pCF10_Tn916 for	TTATCACGCTCGGACTATTGAC

pCF10_Tn916 rev	CGATTCAGAAATTGCAGACCTG
pCF10_uvaB for	GATCGAAATCAGCACATGGAAC
pCF10_uvaB rev	TTTGGCTTATCCGTTTTAACCG
Quantitative RT_PCR primers	
RealTime_11622_For	GCAGAGAATGTGGGAGGACT
RealTime_11622_Rev	TTCTTTCTCGTGCTTGTGGC
RealTime_11823_For	TTTTGCGGTTGCGGTACTGT
RealTime_11823_Rev	TTCGAGGTTCAAGTCGCTTACT
RealTime_10790_For	GAGCAGAACGTACCCCACAA
RealTime_10790_Rev	CCGTTGAGAAAACGCGACTG
RealTime_clpX_For	GAACGTAATACTGGCGCACG
RealTime_clpX_Rev	TTTACCGGTACCTTCAGCGG

Table S2. Extended analysis of SMRT and bisulfite sequencing data.

Table S2A. *E. faecalis* OG1RF sequence variants detected by SMRT sequencing.

Position	Strand	Variant	Type	Coverage	Confidence	Overlapping annotation	Amino acid substitution
23656	+	23656A>G	Substitution	272	93	OG1RF_10019; *manX2*	Thr>Ala
819012	+	819012_819013insT	Insertion	100	93	OG1RF_r_10004; 16S rRNA	
819260	+	819260G>A	Substitution	45	93	OG1RF_r_10004; 16S rRNA	
819271	+	819271C>T	Substitution	44	93	OG1RF_r_10004; 16S rRNA	
1662413	+	1662413C>A	Substitution	120	93	OG1RF_11594; oxidoreductase	Ser>Ile

Table S2B. Analysis of sequences surrounding positions with QV >40 in SMRT sequencing.

Quantity of GCWGC in surrounding context	Number of positions	Mean QV[a]	Std QV[a]	Signal enrichment (Percent occurrence)[b]
0	2077	48.04	66.81	No enriched motif
1	1072	52.41	12.68	N(N_4)GCWGC (15.21%) (163/1072) NNGCWGC (42.72%) (458/1072) GCWGC (17.35%) (186/1072)
2	396	63.64	23.45	N(N_7)GCWGCWGC (11.11%) (44/396) N(N_4)GCWGCWGC (26.26%) (104/396) NNGCWGCWGC (38.64%) (153/396)
3	55	75.82	39.53	GCWGCWGCWGC (58.185) (32/55)
4	10	69.40	20.12	Contain at least one GCWGCWGC
Total (QV>40)	3610	51.53	52.28	

[a]Mean and standard deviation of QVs for modified positions
[b]Percentage occurrences >10% are shown. The underlined positions indicate positions with QV >40.

Table S2C. Assembly statistics for bisulfite sequencing.

	Total reads	# of reads mapped to reference sequence[a]	# of mapped reads after filter[b]	% C positions covered[c]	Mean coverage depth[c]	Mean methylation ratio[d]
OG1RF gDNA	7,255,590	3,893,581	2,927,274	83.7	56.03	0.04286
WGA OG1RF gDNA	1,735,910	617,489	616,682	42.1	20.04	0.01691
BW25113	3,806,930	1,393,975	1392113	63.6	20.40	0.02731
JW1944	3,722,370	1,004,669	1003679	49.6	17.83	0.01275

[a]Paired-end mapping by Bowtie 2
[b]Mapped reads with bisulfite conversion rate <80% were excluded
[c]Conversion rate filter applied; coverage depth threshold applied (≥7X coverage)
[d]Mean methylation ratio for all covered C positions with the conversion rate filter applied

Table S2D. Number of covered cytosines before and after application of conversion rate filter in bisulfite sequencing.

Sample	Total # of cytosine in genome[a]	# of covered C before filter[b]	# of covered C after filter	# of C after filter and with coverage ≥7X
OG1RF	1034280	1023870 (98.99%)	1020912 (98.71%)	865233 (83.66%)
WGA OG1RF	1034280	903219 (87.33%)	889553 (86.00%)	435854 (42.14%)
BW25113	2356477	2297949 (97.52%)	2289786 (97.17%)	1498465 (63.59%)
JW1944	2356477	2242744 (95.17%)	2219752 (94.20%)	1169036 (49.61%)

[a]Number sums C positions from both strand
[b]Percent coverage is shown in parentheses

Table S3. Comparison of motif detection in SMRT and bisulfite (BS) sequencing.

	Motif[a]	Sequencing method	% Motif Detected[b]	% Covered Motif Detected[c]	# Motifs Detected[d]	# Covered Motifs	# of Motifs in Genome	Mean QV[e]	Mean Motif Coverage
OG1RF	G<u>C</u>WGC	BS	83.3%	100%	6981	6981	8380	53.4	54.8
	G<u>C</u>WGC	SMRT	5.1%	5.10%	425	8326	8380	61.6	120.3
	<u>N</u>NGCWGCN<u>N</u>	SMRT	9.5%	9.51%	792	8326	8380	60.8	121.3
	<u>N</u>(N_4)GCWGC(N_4)<u>N</u>	SMRT	4.2%	4.19%	349	8326	8380	59.4	120.2
E. coli BW25113	C<u>C</u>WGG	BS	66.5%	99.99%	16032	16033	24100	60.63	20.03

[a]The predicted modified position is underlined.

[b]Percent motif detection was calculated by dividing the number of modified motifs detected by the total number of motifs in the genome.

[c]Percent covered motif detection was calculated by dividing the number of modified motifs detected by the total number of covered. motifs in the genome with coverage threshold >7X applied.

[d]Number of modified motifs detected.

[e]Mean QV for modified bases in motifs.

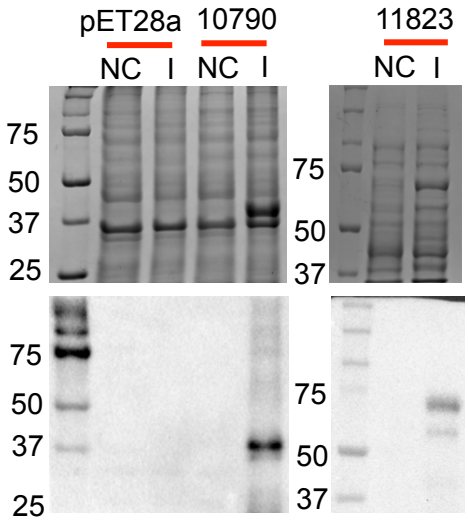

Figure S1. Overexpression of predicted DNA MTases in *E. coli* BL21DE3. *E. coli* expressing predicted DNA MTases were induced overnight using IPTG as described in Material and Methods. The crude cell lysate was analyzed on SDS-PAGE by coomassie blue staining (top panel) and western blot (bottom panel). The predicted molecular weights of the OG1RF_10790 and OG1RF_11823 proteins are 38.0 kDa and 49.7 kDa, respectively. The OG1RF_11823 protein has a GST fusion, which increases its molecular weight by 26 kDa. NC, not IPTG-induced; I, induced with 1mM IPTG.

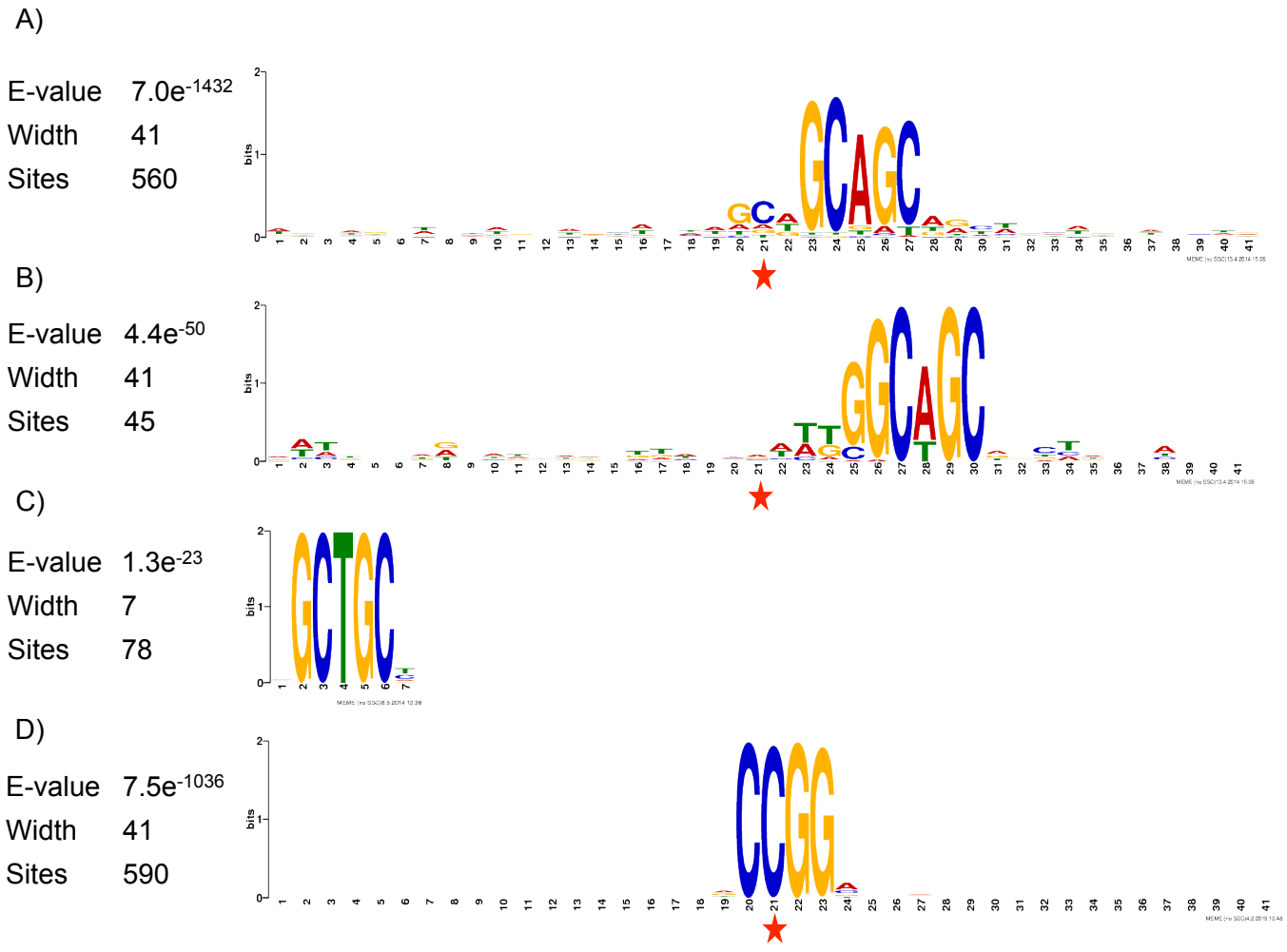

Figure S2. Motif enrichment within the contexts of top 1000 high QV positions from SMRT sequencing. MEME was used to analyze sequence surrounding the 1000 nucleotide positions with the highest QVs. Position 21 is the base with high QV, indicated by a red asterisk.

(A-C) In OG1RF SMRT sequencing, similar motifs were found 2 bases (A) and 5 bases (B) downstream of 560 and 45 high QV positions, respectively, indicating that the high QV positions are actually secondary peaks resulting from modification at GCWGC motifs. Excluding the sequences enriched in (A) and (B), GCTGC motifs with varying positions within the 41 base window were detected by MEME for 78 high QV positions (C).

(D) In *E. coli* BL21DE3 with overexpressed OG1RF_11823 SMRT sequencing, the high QV positions are center enriched at position 21, suggesting a strong primary signal within CCGG motifs. The remaining 410 positions do not share any consensus motif.

Expectation values are shown.


(A)
clpX reference
Bisulfite-converted dCTP PCR product
Bisulfite-converted dm5CTP PCR product
Bisulfite-converted dm4CTP PCR product
Consensus
Coverage
REV 1. clpX pcr pr...
clpX CDS; binds and unfolds substrates as part of the ClpXP protease
clpX gene
REV 2. 142EDAA...
REV 3. 142EDAA...
FWD 4. 142EDAA...
(B)
Reference
BS-converted OG1RF gDNA
(C)
Reference
BS-converted E. coli gDNA

Fig. S3. (A) Bisulfite-converted PCR products distinguish different cytosine methylations. The *E. faecalis* OG1RF *clpX* gene was amplified by PCR using dCTP, dm5CTP, or dm4CTP. Products were bisulfite-converted, re-amplified by PCR with dCTP, and then sequenced. The top line shows the OG1RF *clpX* reference sequence. Sequences of bisulfite-converted products are aligned below the reference sequence. For the dCTP PCR product, all C positions have been converted to T. For the dm5CTP PCR product, no C positions have been converted. For the dm4CTP product, mixed C/T signals are observed for every reference C position.

(B) GCWGC motifs are fully protected from bisulfite conversion. The *E. faecalis* OG1RF_11844 gene contains two GCWGC motifs. The “Reference” line above shows OG1RF_11844 sequence; GCWGC motifs are indicated by red bars below the figure. OG1RF gDNA was bisulfite-converted, OG1RF_11844 sequence was amplified by PCR with dCTP, and products were sequenced. The “BS converted fragment” shows the PCR product sequence aligned to the reference sequence. All C positions were fully converted to T, except for the internal C positions of the two GCWGC motifs. Those C positions were fully protected from bisulfite conversion, suggesting they are modified by m5C.

(C) CCGG motifs are partially protected from bisulfite conversion. The *E. coli* BL21DE3 ECD_00002 gene contains two CCGG motifs. The “Reference” line above shows BL21DE3 sequence; two CCGG motifs are indicated by red bars below the figure. The gDNA from BL21DE3 with overexpressed OG1RF_11823 was bisulfite converted and amplified by PCR with dCTP using primers targeting ECD_00002, and products were sequenced. The “BS converted fragment” shows the sequenced PCR products aligned to reference genome. All C positions were successfully converted to T, except for the internal C of the two CCGG motifs. The mixed population of C and T within CCGG motifs indicates partial protection from bisulfite conversion, suggesting they are modified by m4C.

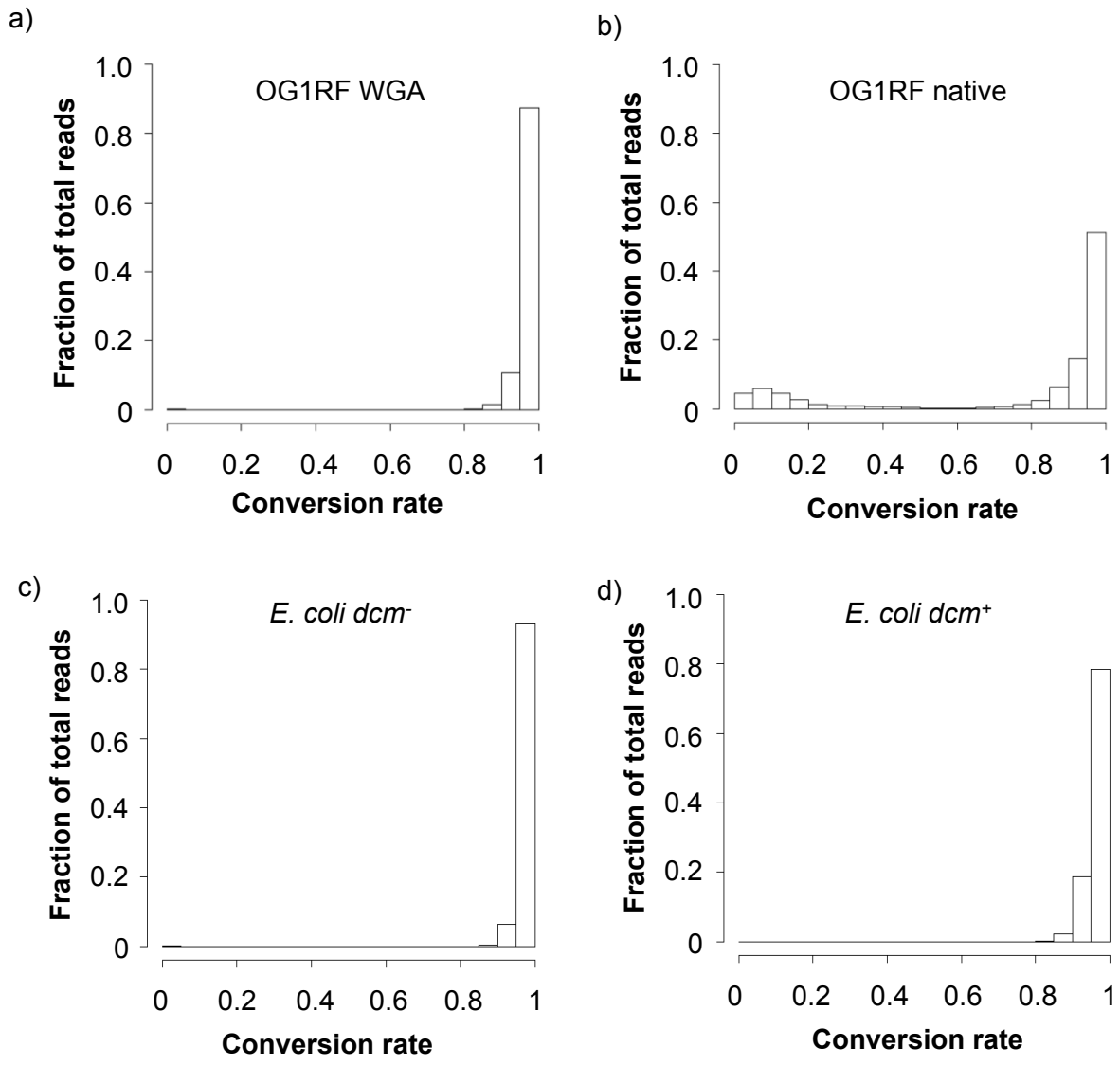

Figure S4. Evaluation of bisulfite conversion rate for bisulfite sequencing reads. Bisulfite conversion rate for each sequence read was calculated as described in the text. Graphs show the distribution of conversion rates for mapped sequencing reads in each sample. (A) OG1RF WGA DNA. (B) OG1RF native gDNA. (C) *E. coli dcm*$^{-}$. (D) *E. coli dcm*$^{+}$.