

# Differential adenine methylation analysis reveals increased variability in 6mA in the absence of methyl-directed mismatch repair

Carl J. Stone,<sup>1</sup> Gwyneth F. Boyer,<sup>2</sup> Megan G. Behringer<sup>1,3</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 13.

**ABSTRACT** Methylated DNA adenines (6mA) are a critical epigenetic modification in bacteria that affect cell processes like replication, stress response, and pathogenesis. While much work has been done characterizing the influence of 6mA on specific loci, very few studies have examined the evolutionary dynamics of 6mA over long time scales. We used third-generation sequencing technology to analyze 6mA methylation across the *Escherichia coli* K-12 substr. MG1655 genome. 6mA levels were consistently high across GATC sites; however, we identified regions where 6mA is decreased, particularly in intergenic regions, especially around the –35 promoter element, and within cryptic prophages and IS elements. We further examined 6mA in WT and methyl-directed mismatch repair-deficient (MMR-) populations after 2,400 generations of experimental evolution. We find that, after evolution, MMR- populations acquire significantly more epimutations resulting in a genome-wide decrease in 6mA methylation. Here, clones from evolved MMR- populations display non-deterministic sets of epimutations, consistent with reduced selection on these modifications. Thus, we show that characterization of 6mA in bacterial populations is complementary to genetic sequencing and informative for molecular evolution.

**IMPORTANCE** Methylation greatly influences the bacterial genome by guiding DNA repair and regulating pathogenic and stress-response phenotypes. But, the rate of epigenetic changes and their consequences on molecular phenotypes are underexplored. Through a detailed characterization of genome-wide adenine methylation in a commonly used laboratory strain of *Escherichia coli*, we reveal that mismatch repair deficient populations experience an increase in epimutations resulting in a genome-wide reduction of 6mA methylation in a manner consistent with genetic drift. Our findings highlight how methylation patterns evolve and the constraints on epigenetic evolution due to post-replicative DNA repair, contributing to a deeper understanding of bacterial genome evolution and how epimutations may introduce semi-permanent variation that can influence adaptation.

**KEYWORDS** experimental evolution, mutation, epigenetics, *Escherichia coli*, N<sup>6</sup>-methyladenine, epimutation

DNA methylation introduces genetic variation in a conservative manner by altering the expression and function of DNA without permanently changing the sequence. Bacterial epigenetics, primarily in the form of DNA base methylation, has an outsized role in the regulation of various cellular processes (1–7). The methylated DNA bases found in bacteria and archaea include N<sup>6</sup>-methyladenine (6mA), 5-methylcytosine (5mC), and N<sup>4</sup>-methylcytosine (4mC), which are added by DNA methyltransferases either as part of a restriction-modification system or by solitary methyltransferases without associated

**Editor** Vaughn S. Cooper, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Address correspondence to Megan G. Behringer, [megan.g.behringer@vanderbilt.edu](mailto:megan.g.behringer@vanderbilt.edu).

The authors declare no conflict of interest.

See the funding table on p. 13.

**Received** 30 May 2023

**Accepted** 22 August 2023

**Published** 5 October 2023

Copyright © 2023 Stone et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

restriction enzymes, called orphan methyltransferases (1, 4, 8–11). In gammaproteobacteria, 6mA modified by the orphan methyltransferase Dam at GATC motifs is the most abundant modified base in the genome (12–15). There are 19,120 of these palindromic GATC motifs in the *Escherichia coli* K-12 genome, or one GATC site on average every ~250 bp (Fig. S1) (16). 6mA influences various processes including, but not limited to, replication timing, transposon mobility, gene expression, virulence, and methyl-directed mismatch repair (MMR) (1, 6, 17–28). Many phenotypes regulated by 6mA involve phase variation, such as the pathogenic fimbriae *pap* in uropathogenic *E. coli* and *std* in *Salmonella enterica* serovar Typhimurium (29–31). In these phase-variable phenotypes, whether gene expression is on or off depends on the methylation status of multiple nearby regulator binding sites (32). Thus, the short-term stability but long-term mutability of DNA methylation facilitates these phase-variable phenotypes by allowing for stable regulation of loci that can rapidly switch with epigenetic changes (33).

The most widespread use of 6mA across the *E. coli* genome is in directing MMR. Methyl-directed MMR GATC sites are nearly all methylated on both strands, except for the brief period after chromosome replication before the daughter DNA strand is methylated by Dam (19, 34). When a mismatch is detected by MMR proteins (MutHLS), the nearby hemimethylated GATC sites are used to direct the MMR protein complex to remove the mismatched base from the daughter strand (19, 23, 35, 36). In *E. coli*, inactivation of any of the MutHLS proteins disrupts MMR and leads to a ~150-fold increased mutation rate (37, 38). Since MMR relies on the presence of hemimethylated GATC sites, alterations to the function of Dam in *E. coli* and other gammaproteobacteria also reduce the effectiveness of MMR (18, 22, 39).

Given the role of MMR in preventing replicative errors from becoming permanent mutations, we hypothesize that maintaining MMR imposes selective pressure on the *E. coli* genome to maintain GATC sites and associated 6mA modifications. While 6mA can have local gene expression effects when GATC sites coincide with regulatory elements and transcription factor binding sites, MMR is likely the most significant use of (and influence on) 6mA genome wide. Thus, we predict that in the absence of MMR, *E. coli* experimentally evolved for thousands of generations will exhibit increased variability in per-site 6mA methylation and a genome-wide decrease in 6mA due to a sustained accumulation of epimutations that do not incur the strongly deleterious fitness effect of causing MMR dysfunction. There has been limited examination of the dynamics of 6mA in evolving bacterial populations (40, 41). Thus, in this study, we sought to investigate how the presence and absence of MMR influence 6mA methylation in evolving bacterial populations over an extended time period. To accomplish this, we created a differential methylation-calling pipeline designed for bacterial 6mA data from third-generation sequencing: CoMMA (Comparison of Microbial Methylated Adenines). Third-generation sequencing platforms (such as PacBio's SMRT sequencing or ONT's Nanopore sequencing) allow for accurate and cost-effective identification of DNA base modifications. Using Nanopore sequencing and CoMMA, we first characterized genome-wide methylation frequency at GATC sites in *E. coli* MG1655. Then, focusing on clones isolated from evolved populations with initially WT and MMR-deficient genetic backgrounds, we characterized how 6mA methylation frequency changed among all 19,120 GATC motifs after 2,000 generations of experimental evolution in batch culture (42). We then compared our findings to mutation accumulation studies measuring spontaneous mutation rates in *E. coli* to determine how MMR constrains genetic and epigenetic evolution at GATC sites (37). Together, our results illustrate the evolvability of 6mA and its potential to contribute to heritable change in bacterial populations.

## RESULTS

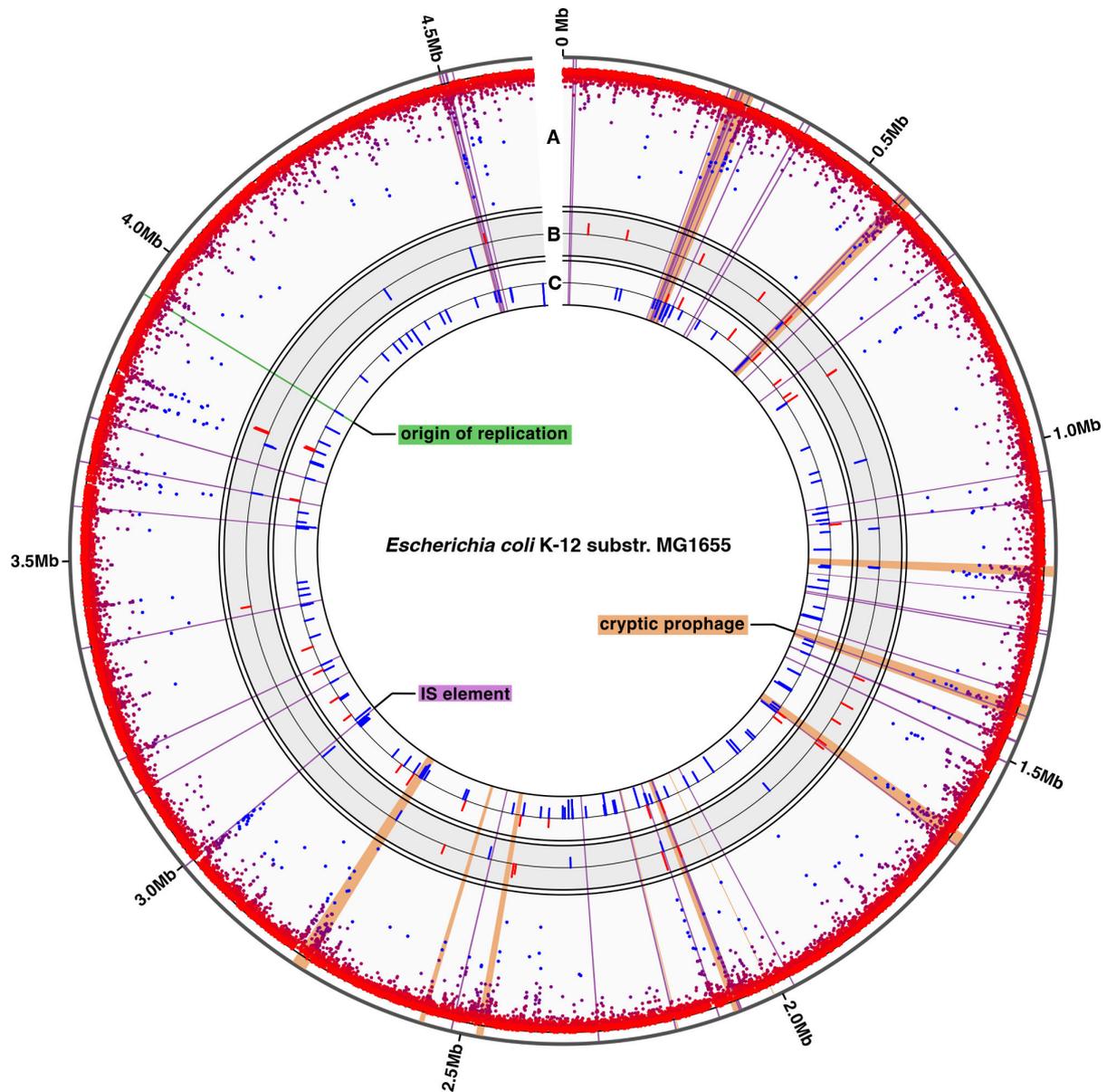
### 6mA methylation status varies between genomic features in *E. coli* MG1655

We first measured genome-wide methylation in *E. coli* K-12 MG1655 (strain PFM2) to characterize methylation trends in our wild-type ancestor (37). Briefly, three colonies from an overnight culture were sequenced on an ONT MinION portable sequencer.

Nanopore sequencing produced a median read depth of 57 reads per GATC site, and there was no significant strand bias in coverage (Fig. S2). To measure methylation from Nanopore reads, we used the methylation classification program Megalodon to calculate the proportion of methylated reads to total reads at each GATC site called the percent methylation (Data set S1) (43). Previous benchmarking studies have shown Megalodon to be one of the most reliable methylation callers when applied to data generated from Nanopore sequencing (44, 45). As there is currently no accepted gold standard methylation calling method, we also compared the performance of multiple independently developed methylation calling programs. With our data, Megalodon most consistently identified the same GATC sites as methylated or hypomethylated (<60% methylated) and correctly identified stably hypomethylated sites that were reported previously in studies that investigated methylation of select loci with biochemical methods (see Materials and Methods, Supplemental Text S1; Figs. S3 through S5; Table S1; Table S2) (34, 46–49). After classifying methylated sites with Megalodon, all GATC sites were annotated with data from RegulonDB, a curated database of *E. coli* K-12 transcriptional regulatory networks (50). Consistent with previous reports, most GATC sites were almost completely methylated, with a median genome-wide percent methylation of 97% (Fig. 1; Fig. S1) (8, 34, 40, 51). Methylation profiling revealed that 177 double stranded GATC sites are hypomethylated on at least one strand, of which 138 were hemimethylated and 39 were hypomethylated on both strands (Table S3). Intergenic GATC sites were significantly less methylated compared to those within coding sequences (median percent methylated reads of 93% compared to 97%; Wilcoxon rank sum test,  $P < 2.2 \times 10^{-16}$ ). This decrease at intergenic sites is likely due to the higher proportion of transcriptional regulatory regions and transcription factor binding sites in intergenic regions, where these sites may exclude binding of the Dam methyltransferase to GATC sites.

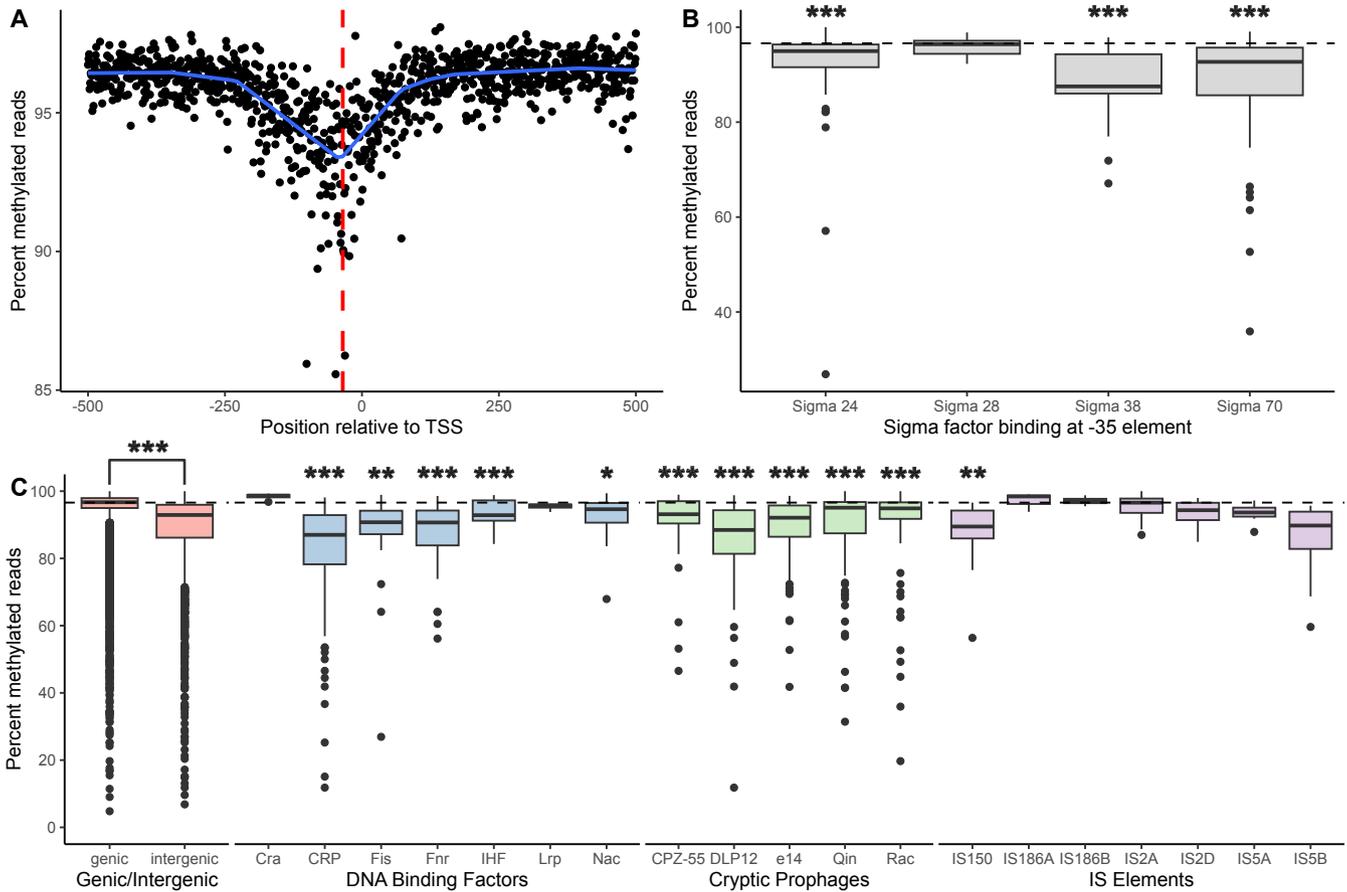
For a broader view of the factors that influence GATC methylation and to determine if hypomethylation was associated with GATC sites that may be shielded by protein binding, we looked at sites within intergenic regions where DNA-protein interactions likely occur, such as sigma factor binding sites and binding sites of other transcriptional regulatory proteins (2, 33, 48, 48, 52). Overall, within these intergenic regions, there is a general decrease in methylation around promoters. GATC sites within –10 and –35 promoter elements were significantly less methylated (95.2% for –10 elements,  $P_{-10} = 6.7 \times 10^{-8}$ ; 90.4% for –35 elements,  $P_{-35} = 6.7 \times 10^{-10}$ , Wilcoxon signed-rank test with Benjamini-Hochberg correction) than the genome-wide median percent methylation (96.6%). This is best observed when methylation is averaged across all promoter regions in the genome; there is a lower frequency of methylated reads around 100 bp before and after the transcription start site, with a maximum decrease in methylation of ~5% centered on the –35 element (Fig. 2A; Fig. S6). Moreover, –35 promoter elements known to be bound by  $\sigma^{70}$ ,  $\sigma^{38}$ , and  $\sigma^{24}$  had significantly lower percent methylation than the genome-wide median (Wilcoxon signed rank test with BH correction:  $P_{\text{sigma}70} = 2.0 \times 10^{-10}$ ;  $P_{\text{sigma}38} = 7.0 \times 10^{-4}$ ;  $P_{\text{sigma}24} = 1.6 \times 10^{-5}$ ) (Fig. 2B).

In addition to core promoter elements, binding sites of other transcriptional regulatory proteins within intergenic regions also exhibited differences in their degree of hypomethylation (Fig. 2C; Fig. S7). Specifically, the global regulators CRP, Fis, Fnr, and Nac were significantly less methylated than the genome-wide median (Wilcoxon signed-rank test with BH correction:  $P_{\text{CRP}} = 2.0 \times 10^{-18}$ ;  $P_{\text{Fis}} = 4.2 \times 10^{-3}$ ;  $P_{\text{Fnr}} = 9.1 \times 10^{-6}$ ;  $P_{\text{Nac}} = 4.98 \times 10^{-2}$ ). Differences in methylation were also observed within coding regions for a few genomic features (Fig. 2C). Across the 12 cryptic prophages in the *E. coli* K-12 genome, significantly decreased methylation is observed in 9 of them: CP4-44, CP4-57, CP4-6, CPS-53, DLP12, e14, Qin, Rac, and PR-Y (CP4-44: 94.0% methylation,  $P = 6.2 \times 10^{-4}$ ; CP4-57: 92.3%,  $P = 9.4 \times 10^{-16}$ ; CP4-6: 95.3%,  $P = 2.5 \times 10^{-13}$ ; CPS-53: 92.4%,  $P = 4.7 \times 10^{-6}$ ; DLP12: 88.5%,  $P = 3.7 \times 10^{-13}$ ; e14: 92.1%,  $P = 1.1 \times 10^{-8}$ ; Qin: 95.1%,  $P = 2.9 \times 10^{-6}$ ; Rac: 94.9%,  $P = 4.3 \times 10^{-8}$ ; PR-Y: 93.8%,  $P = 2.6 \times 10^{-2}$ ; Wilcoxon signed-rank test with BH correction) (53, 54). Of the various IS elements, IS150, IS4, IS30A, IS30C, and IS30D had significantly decreased methylation (IS150: 89.4% methylation,  $P = 2.6 \times 10^{-3}$ ; IS4: 94.1%,



**FIG 1** Genome-wide 6mA methylation in *E. coli* K-12 substr. MG1655. The origin of replication (green), IS elements (purple) and cryptic prophages (orange) are shown for reference. (A) Percent methylation at GATC sites in the ancestral WT MG1655 strain. Each point represents one GATC site and points are colored to reflect percent methylation from 0% (blue, inside) to 100% (red, outside). (B and C) Differentially methylated GATC sites in evolved WT clones (B) and MMR- (C) clones. Red bars are sites with increased methylation and blue bars are decreased methylation, and the height of the bar reflects the difference in percent methylation between the ancestor and the evolved clones. Each ring ranges from  $-25\%$  to  $25\%$ .

$P = 1.3 \times 10^{-2}$ ; IS30A: 89.7%,  $P = 4.3 \times 10^{-2}$ ; IS30C: 85.8%,  $P = 1.5 \times 10^{-2}$ ; IS30D: 86.7%,  $P = 2.2 \times 10^{-2}$  (55–57). Lastly, some genomic features, such as REP elements and terminator sequences, do not show measurable differences in 6mA methylation compared to the genome-wide average. However, these regions either do not contain many GATC sites or are poorly annotated. Comparing sequencing replicates from future studies across a variety of conditions, combined with improving genomic annotations, could allow for the detection of higher-resolution methylation differences and the characterization of methylation on abundant but unannotated sites.



**FIG 2** Genome-averaged methylation decreases around core promoter elements. (A) Median methylation at GATC sites aligned relative to the transcription start sites (TSS) for all promoters. Points represent the median methylation for all GATC sites at that position. The smoothed median was calculated by additive quantile regression smoothing (blue line). The -35 position relative to the TSS is shown for reference (dashed red line). (B) Comparison of percent methylated reads at -35 elements by specific sigma factor binding. Boxplots show the median, first quartile, and third quartile. Sigma factors were compared to the genome-wide median percent methylation (97%, dashed line) with a one-sample Wilcoxon signed-rank test with Benjamini-Hochberg correction. *P*-values are shown in B and C as \*\*\*<0.001, \*\*<0.01, \*<0.05. All groups without stars are not significant. (C) Different genomic features showed variable methylation status. Percent methylation of GATC sites within each feature type was compared to the genome-wide median percent methylation (97%, dashed line) with one-sample Wilcoxon signed-rank tests with Benjamini-Hochberg correction.

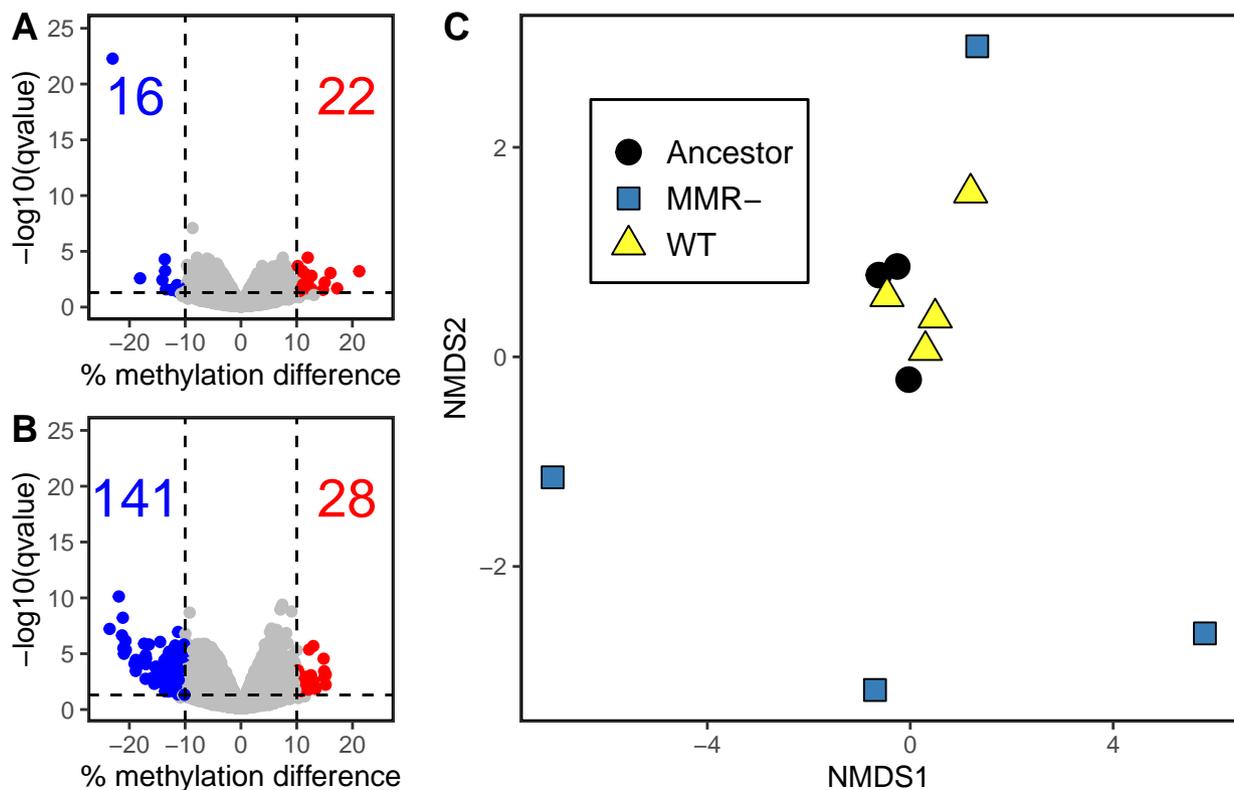
### Differential methylation analysis shows that MMR- clones evolve decreased GATC methylation

After characterizing a baseline *E. coli* K-12 substr. MG1655 6mA methylome, we compared GATC methylation across the experimental ancestor PFM2 strain and eight isolated clones from populations that were experimentally evolved in batch culture for 2400 generations with either intact (WT) or disrupted mismatch repair (MMR-) (42). Briefly, CoMMA adapts the differential methylation package methylKit to identify GATC sites with significant differences in methylation status between bacterial samples using logistic regression (58). For a broad view of how methylation evolves, we focused on two populations from each of the WT or MMR- genetic backgrounds; from each of these populations, we selected two clones and ensured each clone belonged to a different evolved subpopulation. This sampling scheme would allow us to evaluate how methylation evolves within and among populations depending on their ability to conduct methyl-directed mismatch repair.

As with the ancestor, experimentally evolved clones were Nanopore sequenced, and methylated sites were called using Megalodon (Data set S2). The median genome-wide methylation percentage of both genetic backgrounds decreased during batch

culture evolution with MMR- clones exhibiting a significantly larger median decrease in methylation ( $-0.15\%$ ) compared to evolved WT clones ( $-0.03\%$ ; Wilcoxon rank-sum test,  $P < 2.2 \times 10^{-16}$ , Fig. S8). Next, we determined which specific GATC sites were differentially methylated between the evolved clones and the ancestor (Fig. 3A and B). Using cutoffs  $P < 0.05$  and percent methylation change  $>10\%$ , there were significantly more differentially methylated sites in the MMR- clones (28 sites with increased methylation, 141 sites with decreased methylation) compared to the evolved WT clones (22 increased, 16 decreased; Fisher's exact test,  $P < 2.2 \times 10^{-16}$ ).

Both evolved backgrounds also had a significant overrepresentation of differentially methylated sites within certain genomic regions, especially in sites where DNA-protein interactions occur. Across the genome, differentially methylated sites in both WT and MMR- evolved clones appear to coincide with regions of low methylation in the ancestor and with cryptic prophages and IS elements (Fig. 1B and C). Sites that were hemimethylated and hypomethylated in the MG1655 ancestor were enriched for differential methylation in the evolved clones (Fisher's exact test,  $P < 2.2 \times 10^{-16}$ ; Table S3). Differentially methylated sites in both evolved backgrounds were significantly over-enriched in transcription factor binding motifs and IS elements (Fisher's exact test with FDR correction,  $P < 0.05$ , Table 1). Sites with increased methylation in both evolved backgrounds were also over-enriched within cryptic prophages, and sites with decreased methylation were over-enriched in Rho-independent terminators. The MMR- evolved



**FIG 3** Differentially methylated GATC sites between the ancestor and experimentally evolved (A) WT clones or (B) MMR- clones. Differential methylation was called for WT and MMR- clones compared to the ancestor strain using methylKit. Differential methylation was determined for each site as a percent methylation difference compared to the ancestor greater than 10% and a q-value less than 0.05. Numbers shown in each plot field represent the number of differentially methylated sites that increased (red) or decreased (blue) in methylation in the evolved condition compared to the ancestor. (C) Comparison of methylation across all sites between the ancestor and experimentally evolved clones. Euclidean distance between clones was calculated using normalized  $\beta$ -values and ordination was performed with NMDS. Each point represents one clone.

TABLE 1 Overrepresentation of differentially methylated adenines in protein-DNA interaction motifs<sup>a</sup>

Feature type	# GATC sites in ancestor	Evolved WT				Evolved MMR-			
		Increased methylation	<i>P</i> -value	Decreased methylation	<i>P</i> -value	Increased methylation	<i>P</i> -value	Decreased methylation	<i>P</i> -value
Transcriptional regulator									
binding sites	313	7	0.095	12 <sup>c</sup>	< 0.001	11 <sup>b</sup>	0.014	58 <sup>c</sup>	< 0.001
Origin of replication	22	0	1.000	0	1.000	0	1.000	1	0.647
10 element	116	2	0.465	2	0.428	4	0.187	6	0.299
35 element	60	1	0.590	1	0.572	1	0.694	12 <sup>c</sup>	< 0.001
Sigma 24 binding sites	67	2	0.294	1	0.590	2	0.380	4	0.299
Sigma 28 binding sites	11	0	1.000	0	1.000	0	1.000	0	1.000
Sigma 32 binding sites	2	0	1.000	0	1.000	0	1.000	0	1.000
Sigma 38 binding sites	17	0	1.000	0	1.000	0	1.000	3 <sup>c</sup>	0.043
Sigma 70 binding sites	86	1	0.694	2	0.313	3	0.254	12 <sup>c</sup>	< 0.001
Genes	35,595	312	1.000	275	1.000	442	1.000	952	1.000
Rho-dependent terminators	8	0	1.000	0	1.000	0	1.000	1	0.380
Rho-independent terminators	46	1	0.514	4 <sup>c</sup>	0.003	0	1.000	6 <sup>c</sup>	0.011
IS elements	283	10 <sup>b</sup>	0.002	10 <sup>c</sup>	0.001	11 <sup>b</sup>	0.007	18 <sup>c</sup>	0.014
Cryptic prophages	795	15 <sup>b</sup>	0.038	11	0.249	24 <sup>b</sup>	0.001	62 <sup>c</sup>	< 0.001
No feature	1,191	35 <sup>b</sup>	< 0.001	30 <sup>c</sup>	< 0.001	44 <sup>b</sup>	< 0.001	134 <sup>c</sup>	< 0.001

<sup>a</sup>*P*-values were calculated by one-tailed Fisher's exact test followed by BH correction.

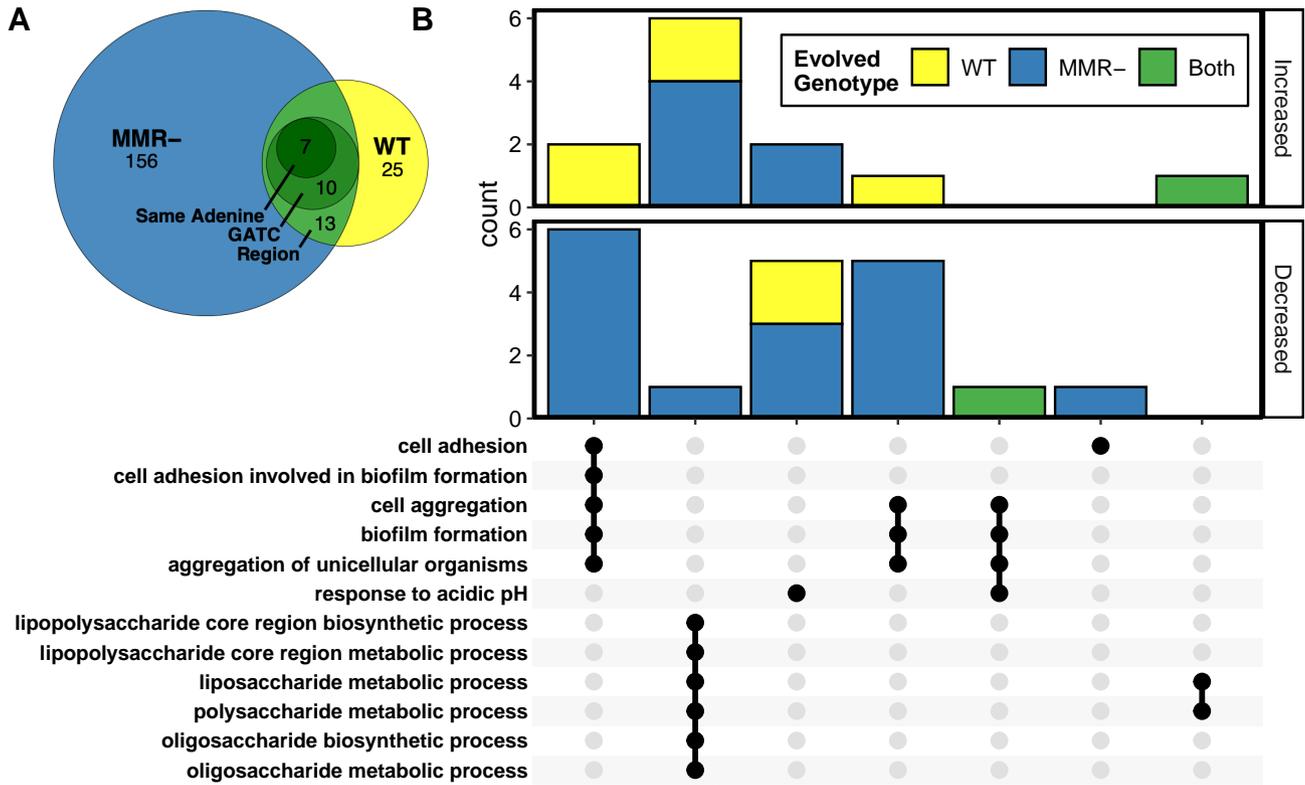
<sup>b</sup>Significant number of sites with increased methylation.

<sup>c</sup>Significant number of sites with decreased methylation.

clones also showed enrichment for under-methylated sites within cryptic prophages ( $P < 2.2 \times 10^{-16}$ ) and within promoters, specifically  $\sigma^{70}$ -binding sites in the  $-35$  region ( $P < 2.2 \times 10^{-16}$ ).

Gene ontology (GO) enrichment analysis of genes containing or near differentially methylated sites revealed enrichment in genes associated with stress response, biofilm formation, and virulence factor production. Of the 194 differentially methylated sites identified between both WT and MMR- evolved clones, 13 sites were in the same gene or intergenic region between the two genotypes (Fig. 4A). Ten of these sites were in the same palindromic GATC site, and 7 of those were in the same adenine residue. GO enrichment analysis was then done on genes that contained differentially methylated GATC sites, or if the GATC site was not within a gene then the nearest gene was used. Genes that were less methylated in the MMR- clones were significantly enriched for biofilm formation and transcriptional regulators (Fig. 4B). Genes that were more methylated after evolution were enriched for biofilm formation in the evolved WT clones and for lipopolysaccharide metabolism in both WT and MMR- clones. Interestingly, 6 of the 15 biofilm genes identified belong to 3 different putative chaperone-usher fimbrial operons. These fimbriae are cryptic in *E. coli* K-12 under laboratory conditions, but when overexpressed they increase biofilm formation (59). Other biofilm genes *yfaL* and *ycgV* are homologs of the cryptic prophage-encoded autotransporter and adhesin Antigen 43 (Ag43 encoded by *flu*) (60). Both fimbriae and Ag43 are known to be subject to methylation-regulated phase variation in *E. coli* and related species (31, 60–62). Other enriched gene ontologies include lipopolysaccharide synthesis (including the non-functional O-antigen polymerase *wbbH*), acid stress response genes, phage defense genes, and various transcriptional regulators involved in stress response (Table S4) (63–66).

Comparison of the full 6mA methylomes between evolved clones and their ancestor revealed a non-deterministic change in 6mA in the absence of MMR consistent with random genetic drift. The percent of methylated reads at all GATC sites was compared between samples by visualizing using non-metric multidimensional scaling (NMDS; Fig. 3C). The evolved WT clones and the ancestor clones clustered together, showing that their 6mA methylomes remain similar after extended evolution (Pairwise PERMANOVA Ancestor vs. WT:  $P_{\text{Ancestor-WT}} = 0.168$ ; PERMDISP2, ANOVA with Tukey's HSD:  $P_{\text{Ancestor-WT}} = 0.99$ ). In contrast, MMR- evolved clones diverge from the ancestor in random directions,



**FIG 4** Gene ontology enrichment analysis of differentially methylated GATC sites. (A) Intersect between differentially methylated sites in evolved WT and MMR- genotypes. Sites in the same region are located in the same gene or the same intergenic region. (B) UpSet plot of significantly enriched gene ontology terms of genes associated with differentially methylated GATC sites that increased or decreased in methylation in evolved WT, MMR-, or both. Bar heights are the number of genes associated with the intersecting GO terms shown below.

and the MMR- clones do not cluster together. The centroid of the MMR- clones was near that of the ancestor and WT clones but significantly different than the WT clones, and MMR- clone dispersions were significantly greater than both the ancestor and WT groups (Pairwise PERMANOVA Ancestor vs MMR-:  $P_{\text{Ancestor-MMR}} = 0.195$ , WT vs MMR-:  $P_{\text{WT-MMR}} = 0.026$ ; PERMDISP2, ANOVA with Tukey's HSD:  $P_{\text{Ancestor-MMR}} = 0.018$ ,  $P_{\text{WT-MMR}} = 0.010$ ). This suggests that 6mA differences that occur during evolution in MMR- lines are non-deterministic, resulting in significant changes to methylation for each sample but not strongly in any one direction for the whole group. MMR- clones from the same population did not cluster together, indicating that shared evolutionary history does not have a large effect on broad 6mA differences over this time scale. Together, the significant number of under-methylated sites and the random divergence of differentially methylated sites in the MMR- clones compared to the ancestor suggest that 6mA is shaped by genetic drift in the absence of MMR.

### MMR- evolved lines accumulate GATC mutations at a rate that indicates relaxed selection

After identifying decreased maintenance of 6mA in the absence of MMR, we sought to genomically identify if the evolution of GATC sites in MMR- clones exhibited patterns of decreased evolutionary constraint. Here, we characterized the fraction of mutations that altered GATC sites during batch culture evolution across clones from both genetic backgrounds. After ~2,400 generations, MMR- clones had acquired a total of 959 mutations with 34 mutations (3.5% of mutations) causing a gain or loss of a GATC

TABLE 2 Number of mutations in GATC sites in evolved clones

Background	Population <sup>a</sup>	Clones <sup>b</sup>	Mutations	GATC gain/loss	Percent mutations in GATC sites	
					(this study)	(37)
MMR	113		257	13	5.1	
		113-1	121	3	2.5	
	113-2	96	3	3.1		
	126		46	1	2.2	
		126-2	313	11	3.5	
		126-4	126	3	2.4	
Total		959	34	3.5	(2.8–4.0) <sup>c</sup>	3.5 (2.8–4.3) <sup>c</sup>
WT	125		4	0	0	
		125-6	1	0	0	
		125-7	7	0	0	
	129		4	0	0	
		129-1	8	0	0	
		129-3	9	0	0	
Total		33	0	0	(0.0–0.0) <sup>c</sup>	4.3 (2.5–6.7) <sup>c</sup>

<sup>a</sup>Mutations shared between clones and occurred before subpopulation divergence.

<sup>b</sup>Mutations unique to clones and occurred after subpopulation divergence.

<sup>c</sup>Bootstrap 95% confidence intervals.

site (Table 2; Table S4). In contrast, the WT clones accumulated a total of 33 mutations, none of which had affected GATC sites. Since the relative lack of mutations in WT clones limited our statistical power, we compared the mutation rates of the clones that were evolved in batch culture to mutation rates calculated in a mutation accumulation (MA) experiment by Lee et al. (37). As MA experiments are conducted using culture conditions that minimize the effects of selection, comparing the GATC mutation rates of clones evolved in batch culture to the GATC mutation rates of MA lines allows us to determine how batch culture GATC mutation rates compare to the neutral expectation. Moreover, as Lee et al. (37) measured mutation rates by conducting MA with the same WT and MMR-PFM2 strain, these mutation rates directly reflect the expected rate of GATC mutation for the WT and MMR- clones if their GATC tetranucleotides are evolving under relaxed selection.

Reanalysis of the MA data to calculate mutation rates in tetranucleotide contexts revealed similar spontaneous GATC mutation rates for WT and MMR- MA lines (Table 2). As such, GATC sites should evolve at the same rate for both genetic backgrounds during batch culture evolution unless selection pressures on GATC sites are different based on the presence or absence of MMR. Notably, during MA, 3.5% of mutations in MMR- lines affect GATC sites. This percentage is equal to the fraction of mutations affecting GATC sites in the MMR- clones and is a stark contrast to the absence of GATC mutations in WT clones during evolution in batch culture. This similarity in the GATC mutation rates from the batch culture evolution and MA experiments suggests that batch culture MMR- clones acquire mutations in GATC sites at the neutral rate and that in the absence of functional MMR, GATC sites may be evolving under relaxed selection. Additional evidence that GATC sites and associated adenine methylation may be evolving under relaxed selection in MMR- clones include that the clones isolated from the MMR- population 113 both acquired mutations resulting in a non-synonymous D20A substitution proximal to the GATC recognition domain of *dam* methyltransferase (67) during evolution in batch culture. While the exact consequences of this mutation have not been examined, this substitution could affect DNA binding and GATC recognition, indicating that perfect *Dam* function may not be necessary in the absence of MMR. Together, these data suggest that MMR is a major factor in maintaining GATC sites and *dam* methyltransferase, and in the absence of MMR these loci are vulnerable to mutation.

## DISCUSSION

In this study, we present what is to our knowledge a first-of-its-kind high-resolution characterization of the *E. coli* K-12 MG1655 6mA methylome that includes a functional and genic classification of 6mA methylation. We identified patterns of reduced methylation at intergenic regions, promoter elements, DNA binding sites, and cryptic prophages. This reduced methylation could result from the physical exclusion of Dam methyltransferase at these sites and potentially cause downstream effects on gene regulation and stress response. The ability to acquire high-resolution methylation data opens the door for future investigations on the specific effects of these methylation differences when paired with transcriptomic and biochemical experiments.

We also created and applied the CoMMA pipeline to investigate the effect of 2,400 generations of evolution on WT and MMR- *E. coli* MG1655 clones. We found that the MMR- clones exhibited a significant decrease in methylation, indicating that 6mA is somewhat dispensable in the absence of MMR. While methylation was largely stable at the majority of GATC sites, MMR- clones exhibited significantly more differentially methylated sites than WT clones. Interestingly, these differentially methylated sites are often located in DNA binding sites. Adenine modification can affect specific binding affinity where 6mA coincides with a DNA binding site (68, 69). Without MMR imposing a genome-wide pressure to maintain 6mA for the purpose of directing DNA repair machinery, MMR- clones are free to alter 6mA at these sites with a bias toward decreased methylation, likely due to binding competition between transcription factors and the methyltransferase. Altered 6mA methylation that affects transcription factor binding can result in downstream effects and further alter cellular phenotypes. Our study demonstrates that differential methylomics is a powerful approach that can directly resolve differences between clones that are not apparent from genetic changes alone. Thus, the CoMMA pipeline is a useful tool that can complement genetic approaches.

An interesting observation from our study involves one of the MMR- evolved populations (P113), which acquired a mutation in the *dam* adenine methyltransferase resulting in a single amino acid substitution (DamD20A). The consequence of this mutation is unclear, although this residue is proximal to three S-adenosylmethionine-binding sites at K10, K14, and D54, and could potentially affect protein folding (67). While there are no broad methylation differences within these clones compared to others of the same background, a DamD20A substitution may have contributed to the observed changes in methylation in these clones. Further studies are needed to investigate if this mutation altered Dam activity and how this affects genome-wide methylation patterns.

One potential impact of our findings is how 6mA methylation may influence the mutational spectrum and adaptive landscape of bacteria. DNA base modifications can affect the rates of spontaneous deamination; for instance, 5mC is vulnerable to deamination and contributes to the C→T mutation rate (70–72). Similarly, adenine and 6mA may deaminate to hypoxanthine, which pairs with cytosine during replication (73, 74). While the spontaneous deamination rates of adenine versus 6mA have not been exhaustively studied, recent work identified a positive correlation between 6mA and A→G transitions in *Neisseria meningitidis* (75). Although we did not detect any immediate differences in the mutation spectrum of WT and MMR- MA lines at GATC sites, genome-wide decreases in 6mA after MMR loss may eventually result in a decreased A→G transition rate. This change to the mutational spectrum could potentially affect how MMR- or Dam- cells adapt to changing environments and may influence the trajectory of their evolutionary paths.

This study is the first-ever analysis of differential methylation between laboratory evolved bacterial lineages at this time scale. The ability to identify changes to 6mA at GATC sites over evolutionary time creates opportunities for the seamless integration of genetic and epigenetic data. Nanopore sequencing and pipelines like CoMMA are sufficiently reliable, affordable, and high-throughput to powerfully analyze genetic and epigenetic evolution in bacteria. However, many factors can influence methylation studies, including sequencing chemistry, the methylation calling program, the biological

conditions of the sample, and bioinformatic filtering, data processing, and statistical analysis; thus, care must be taken to identify and account for these systematic biases in comparative studies. The development of standardized tools for methylation analyses like CoMMA is a step toward addressing some of these biases. As more studies uncover how bacterial DNA methylation affects gene expression, differential epigenetic analyses can be extended through time or between bacterial isolates and help to elucidate the relationship between mutation, epimutation, and gene expression as it relates to adaptation to challenging environments (40, 41, 76, 77).

## MATERIALS AND METHODS

### Bacterial strains and culture conditions

All bacterial strains were cultured and sequenced to identify mutations as part of a long-term batch culture experimental evolution study that has been described previously (42). Briefly, the WT ancestor strain PFM2 is a prototrophic derivative of *E. coli* K-12 MG1655, and the MMR-deficient strain PFM5 is directly derived from PFM2 by introducing an in-frame deletion of the *mutL* gene (37, 42). Experimental evolution in batch culture was started by inoculating 10 mL of LB-Miller broth (BD Difco) in 16 × 100 mM glass culture tubes with a single colony of the progenitor strain (42). Cultures were maintained shaking at 175 rpm, alternating transfers after 24 hours of growth at 37°C and 48 hours of growth at 25°C. Cultures of interest in this study were transferred by thoroughly vortexing the culture tube and inoculating 9 mL of LB-Miller broth with 1 mL of culture, resulting in a transfer of about 10<sup>9</sup> cells (42). Populations were grown and transferred for 2,000 total generations of growth (2.5 years). After 2,000 generations, samples from the evolving populations were streaked for isolation and eight random clones were saved in 40% glycerol at –80°C for further and future experimentation.

### Nucleic acid isolation and Nanopore sequencing

Evolved and WT ancestor clones were revived from frozen storage on LB agar by incubating overnight at 37°C. For each clone, a single isolated colony was randomly selected and restreaked on to fresh LB agar and again incubated overnight at 37°C for DNA extraction. Following overnight incubation, colonies were collected from the agar using a 10 mL loop and inoculated directly into DNA extraction buffer, and DNA was further isolated with the UltraClean Microbial DNA Kit (Qiagen). Following recommendations from ONT, extracted DNA was assessed for quantity using the Qubit High Sensitivity dsDNA kit and quality using the Genomic DNA ScreenTape on a TapeStation 4200 (Agilent). Libraries were then prepared for Nanopore sequencing using the PCR-free Rapid DNA Sequencing Kit (SQK-RAD004, Nanopore) paired with the Rapid Barcoding Kit (SQK-RBK004). Prepared libraries were pooled and loaded on to a R.9.4.1 flow cell and sequenced on a MinION for 48 hours with live basecalling in MinKNOW deactivated.

### Bioinformatics and sequencing analysis

To determine the precision of methylation calling, we compared results for each of four methylation calling pipelines across three Nanopore sequencing replicates of the WT ancestor PFM2 and determined the number of inconsistently called sites. Initial Nanopore basecalling by converting fast5 files into fastq was conducted using Guppy GPU (v. 3.1.5+781ed57). Following basecalling, samples were demultiplexed with Guppy GPU. Average read length across all samples is 5,583 ± 275 nucleotides (± s.e.m). Following basecalling and demultiplexing, we evaluated four different pipelines [Tombo (v.1.5), mCaller (v.0.3), DeepSignal (v.0.1.6), and Megalodon (v.0.1.0)] for their precision and accuracy to identify base modifications and determine methylated adenines within GATC sites (43, 46, 47, 78). Both Tombo and Megalodon were created by ONT for methylation calling, with Tombo being an older release that relies on statistical methods to annotate non-canonical bases and Megalodon being a newer release based on

machine learning with deep neural networks (43, 47). To increase accuracy, Megalodon can be run with a VCF file containing the position and identity of known variants, for which we used Illumina sequencing data for evolved clones that was previously published (42). For pipelines not developed by ONT: DeepSignal applies a deep neural network to data that has been preprocessed by Tombo; and mCaller applies a deep neural network to data that has been preprocessed by Nanopolish. When running the mCaller pipeline we used Nanopolish v.0.11.1). Across all samples, the mean read coverage across GATC sites was 57 reads per site (Fig. S2).

Identification of mutations affecting GATC sites in evolved clones was conducted using previously published WGS Illumina sequencing data. To identify the rate of mutations affecting GATC sites under minimal selection, we utilized previously published mutation accumulation data for the *E. coli* strains that served as the experimental ancestor for the WT (PFM2) and MMR- (PFM5, PFM2: $\Delta$ mutL) (37). For both the experimentally evolved clones and the MA lines, mutations were called using GATK best practices and the genic location of mutations was annotated using SnpEff v.4.3 by comparing VCF files of identified mutations for each clone to the *Escherichia coli* K-12 str. MG1655 reference genome (INSDC accession: U00096.3) (79). To confirm a mutation occurred in a GATC site, the tetranucleotide context of each mutation was identified with the fill-fs -l 5 function from vcftools (v.0.1.12) to annotate each variant with the flanking five upstream and downstream nucleotides (80).

GATC sites were annotated as within different genomic features using all combined databases in RegulonDB v.10.9 (81). These annotations were concatenated into a BED file, which was used to annotate GATC sites with the CoMMA function `annotateMethylSites`.

### Differential methylation calling

Differential methylation analysis was performed in R using CoMMA, a package that serves as a wrapper for methylKit v.3.13 (58). GATC sites that had acquired mutations in any clone were removed from differential methylation analysis so that only GATC sites shared between samples were considered. To ensure the reliability of differential methylation analysis, CoMMA includes the following modifications to the standard methylKit workflow: (i) any sites covered by fewer than 10 reads were excluded from analysis; (ii) methylation coverage at each site was normalized by median coverage; and (iii) sites were ultimately considered to be differentially methylated based on a q-value cutoff of <0.05 and percent methylation difference >10%. Differential methylation was called separately between the ancestor and either the evolved MMR- or WT clones. Differential methylation was calculated by logistic regression with a sliding linear model (SLIM) to correct for multiple hypothesis testing (58, 82).

### Statistical analysis and ordination

Statistical analysis comparing percent methylation (or  $\beta$ -value) was performed on M-values to remove the heteroscedasticity observed in  $\beta$ -value distributions (83). M-values were calculated as shown in Equation 1:

$$M = \log_2\left(\frac{\beta}{1 - \beta + 0.001}\right) \quad (1)$$

All statistical analysis was done in R. M-values of GATC sites within different genomic features were compared using the Kruskal-Wallis rank sum test, and each feature was compared to the genome-wide median M-value using a one-sample Wilcoxon signed-rank test with Benjamini-Hochberg correction for multiple comparisons (84).

Ordination of samples was performed after normalizing  $\beta$ -values. Since  $\beta$ -values are dependent on coverage, samples with low coverage stand out when visualized with non-metric ordination methods. Therefore, before ordination, we normalized coverage and methylated read counts between samples to reduce the effects of coverage on ordination distance. First, the coverage of all samples was normalized using quantile

normalization (85). Then, a scaling equation was calculated for each sample using the simple linear regression line between coverage and normalized coverage. This scaling equation was then applied to each sample's coverage and methylated read count, both of which were used to calculate normalized  $\beta$ -values. Ordination of samples was performed using the R package *vegan* v.2.6–4 (86). Non-metric multidimensional scaling was used to visualize sample clustering based on a Euclidean distance matrix calculated from normalized  $\beta$ -values. PERMANOVA and PERMDISP2 were done with *vegan* functions *adonis2* and *betadisper*, respectively.

Gene ontology enrichment analysis was performed using R package *clusterProfiler* v.4.6.2 and the Bioconductor *E. coli* K-12 annotation v.3.8.2 (87–89). Genes enriched for biological process ontologies were filtered by a q-value cutoff of 0.05, and redundant GO terms were removed using the *clusterProfiler* “simplify” method.

## ACKNOWLEDGMENTS

We would like to thank W.-C. Ho, H. Lee, M. Lynch, and K. Samerotte for their helpful comments. High-performance computing resources were provided by the Research Computing Core at Arizona State University and the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.

This work was supported by Army Research Office grants W911NF-21-1-0161 (M.G.B.), National Institutes of Health grant F32GM123703 (M.G.B.), and start-up funding from Vanderbilt University and the Vanderbilt Center for Infection, Immunology, and Inflammation.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

<sup>2</sup>Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

<sup>3</sup>Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

## AUTHOR ORCIDs

Megan G. Behringer  <http://orcid.org/0000-0001-7036-6014>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">HHS   NIH   National Institute of General Medical Sciences (NIGMS)</a>	F32GM123703	Megan G. Behringer
<a href="#">DOD   USA   AFC   CDC   ARO   Life Sciences Division, Army Research Office (Life Sciences Division ARO)</a>	W911NF-21-1-0161	Megan G. Behringer

## AUTHOR CONTRIBUTIONS

Carl J. Stone, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Gwyneth F. Boyer, Formal analysis, Investigation, Methodology, Resources, Writing – original draft | Megan G. Behringer, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

## DATA AVAILABILITY

Newly generated sequencing data for the identification of nucleotide modifications can be downloaded from SRA, BioProject [PRJNA912686](https://ncbi.nlm.nih.gov/bioproject/PRJNA912686). Code and reformatted data for reproducing this analysis are found at <https://github.com/BehringerLab/Methylation>. The CoMMA R package is available at <https://github.com/carl-stone/CoMMA>.

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Data Set S1 (mBio01289-23-s0001.xlsx)**. Methylation data for the three ancestor replicates called with all 4 methylation calling software packages.

**Data Set S2 (mBio01289-23-s0002.xlsx)**. Methylation data for the Ancestor and Evolved clones called with Megalodon.

**Text S1 (mBio01289-23-s0003.pdf)**. Evaluation of methylation calling software packages.

**Supplemental Figures (mBio01289-23-s0004.pdf)**. Figures S1 to S8.

**Table S1 (mBio01289-23-s0005.xlsx)**. Classification of methylation status at double-stranded GATC sites for each methylation calling program.

**Table S2 (mBio01289-23-s0006.xlsx)**. Proportion of methylated reads at nonmethylated GATC sites across all callers.

**Table S3 (mBio01289-23-s0007.xlsx)**. All hemimethylated and hypomethylated GATC sites in the WT K-12 MG1655 genome.

**Table S4 (mBio01289-23-s0008.xlsx)**. GATC sites lost or gained by mutation in MMR-evolved clones.

## REFERENCES

- Sánchez-Romero MA, Casadesús J. 2020. The bacterial epigenome. *Nat Rev Microbiol* 18:7–20. <https://doi.org/10.1038/s41579-019-0286-2>
- Wion D, Casadesús J. 2006. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol* 4:183–192. <https://doi.org/10.1038/nrmicro1350>
- Seib KL, Srikhanta YN, Attack JM, Jennings MP. 2020. Epigenetic regulation of virulence and immunoevasion by phase-variable restriction-modification systems in bacterial pathogens. *Annu Rev Microbiol* 74:655–671. <https://doi.org/10.1146/annurev-micro-090817-062346>
- Adhikari S, Curtis PD. 2016. DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol Rev* 40:575–591. <https://doi.org/10.1093/femsre/fuw023>
- Ardissone S, Redder P, Russo G, Frandi A, Fumeaux C, Patrignani A, Schlapbach R, Falquet L, Viollier PH. 2016. Cell cycle constraints and environmental control of local DNA hypomethylation in  $\alpha$ -proteobacteria. *PLoS Genet* 12:e1006499. <https://doi.org/10.1371/journal.pgen.1006499>
- Reisenauer A, Kahng LS, McCollum S, Shapiro L. 1999. Bacterial DNA methylation: a cell cycle regulator *J Bacteriol* 181:5135–5139. <https://doi.org/10.1128/JB.181.17.5135-5139.1999>
- Collier J, McAdams HH, Shapiro L. 2007. A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl Acad Sci U S A* 104:17111–17116. <https://doi.org/10.1073/pnas.0708112104>
- Sánchez-Romero MA, Cota I, Casadesús J. 2015. DNA methylation in bacteria: from the methyl group to the methylome. *Curr Opin Microbiol* 25:9–16. <https://doi.org/10.1016/j.mib.2015.03.004>
- De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, Haigh R, Croucher NJ, Oggioni MR. 2017. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. *FEMS Microbiol Rev* 41:S3–S15. <https://doi.org/10.1093/femsre/fux025>
- Nye TM, Jacob KM, Holley EK, Nevarez JM, Dawid S, Simmons LA, Watson ME. 2019. DNA methylation from a type I restriction modification system influences gene expression and virulence in *Streptococcus pyogenes*. *PLoS Pathog* 15:e1007841. <https://doi.org/10.1371/journal.ppat.1007841>
- Oliveira PH, Fang G. 2021. Conserved DNA methyltransferases: a window into fundamental mechanisms of epigenetic regulation in bacteria. *Trends Microbiol* 29:28–40. <https://doi.org/10.1016/j.tim.2020.04.007>
- Geier GE, Modrich P. 1979. Recognition sequence of the *dam* methylase of *Escherichia coli* K12 and mode of cleavage of DpnI endonuclease. *J Biol Chem* 254:1408–1413.
- Herman GE, Modrich P. 1982. *Escherichia coli dam* methylase. physical and catalytic properties of the homogeneous enzyme. *J Biol Chem* 257:2605–2612.
- Marinus MG, Morris NR. 1973. Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *J Bacteriol* 114:1143–1150. <https://doi.org/10.1128/jb.114.3.1143-1150.1973>
- Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, Posfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM, Korlach J, Pennacchio LA, Roberts RJ. 2016. Epigenomic landscape of prokaryotes. *PLoS Genet* 12:e1005854. <https://doi.org/10.1371/journal.pgen.1005854>
- Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1462. <https://doi.org/10.1126/science.277.5331.1453>
- Boye E, Løbner-Olesen A. 1990. The role of *dam* methyltransferase in the control of DNA replication in *E. coli*. *Cell* 62:981–989. [https://doi.org/10.1016/0092-8674\(90\)90272-g](https://doi.org/10.1016/0092-8674(90)90272-g)
- Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M. 1983. Effects of high levels of DNA Adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics* 104:571–582. <https://doi.org/10.1093/genetics/104.4.571>
- Schlagman SL, Hattman S, Marinus MG. 1986. Direct role of the *Escherichia coli dam* DNA methyltransferase in methylation-directed

- mismatch repair. *J Bacteriol* 165:896–900. <https://doi.org/10.1128/jb.165.3.896-900.1986>
20. Messer W, Noyer-Weidner M. 1988. Timing and targeting: the biological functions of dam methylation in *E. coli*. *Cell* 54:735–737. [https://doi.org/10.1016/s0092-8674\(88\)90911-7](https://doi.org/10.1016/s0092-8674(88)90911-7)
  21. Messer W, Bellekes U, Lother H. 1985. Effect of *dam* methylation on the activity of the *E. coli* replication origin, *oriC*. *EMBO J* 4:1327–1332. <https://doi.org/10.1002/j.1460-2075.1985.tb03780.x>
  22. Herman GE, Modrich P. 1981. *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. *J Bacteriol* 145:644–646. <https://doi.org/10.1128/jb.145.1.644-646.1981>
  23. Lahue RS, Modrich P. 1988. Methyl-directed DNA mismatch repair in *Escherichia coli*. *Mutat Res* 198:37–43. [https://doi.org/10.1016/0027-5107\(88\)90037-1](https://doi.org/10.1016/0027-5107(88)90037-1)
  24. Roberts D, Hoopes BC, McClure WR, Kleckner N. 1985. IS10 transposition is regulated by DNA adenine methylation. *Cell* 43:117–130. [https://doi.org/10.1016/0092-8674\(85\)90017-0](https://doi.org/10.1016/0092-8674(85)90017-0)
  25. Tomcsanyi T, Berg DE. 1989. Transposition effect of adenine (Dam) methylation on activity of O end mutants of IS50. *J Mol Biol* 209:191–193. [https://doi.org/10.1016/0022-2836\(89\)90271-4](https://doi.org/10.1016/0022-2836(89)90271-4)
  26. Torreblanca J, Casadesús J. 1996. DNA adenine methylase mutants of *Salmonella typhimurium* and a novel *dam*-regulated locus. *Genetics* 144:15–26. <https://doi.org/10.1093/genetics/144.1.15>
  27. Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ. 1999. An essential role for DNA adenine methylation in bacterial virulence. *Science* 284:967–970. <https://doi.org/10.1126/science.284.5416.967>
  28. Lu M, Campbell JL, Boye E, Kleckner N. 1994. SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77:413–426. [https://doi.org/10.1016/0092-8674\(94\)90156-2](https://doi.org/10.1016/0092-8674(94)90156-2)
  29. Blyn LB, Braaten BA, Low DA. 1990. Regulation of *pap* pilin phase variation by a mechanism involving differential *dam* methylation States. *EMBO J* 9:4045–4054. <https://doi.org/10.1002/j.1460-2075.1990.tb07626.x>
  30. Jakomin M, Chessa D, Bäumlner AJ, Casadesús J. 2008. Regulation of the *Salmonella enterica* *std* fimbrial operon by DNA adenine methylation, SeqA, and HdfR. *J Bacteriol* 190:7406–7413. <https://doi.org/10.1128/JB.01136-08>
  31. Hernday AD, Braaten BA, Low DA. 2003. The mechanism by which DNA adenine methylase and PapI activate the Pap epigenetic switch. *Mol Cell* 12:947–957. [https://doi.org/10.1016/s1097-2765\(03\)00383-6](https://doi.org/10.1016/s1097-2765(03)00383-6)
  32. Van der Woude MW. 2011. Phase variation: how to create and coordinate population diversity. *Curr Opin Microbiol* 14:205–211. <https://doi.org/10.1016/j.mib.2011.01.002>
  33. Sánchez-Romero MA, Casadesús J. 2021. Waddington's landscapes in the bacterial world. *Front Microbiol* 12:685080. <https://doi.org/10.3389/fmicb.2021.685080>
  34. Tavazoie S, Church GM. 1998. Quantitative whole-genome analysis of DNA-protein interactions by *in vivo* methylase protection in *E. coli*. *Nat Biotechnol* 16:566–571. <https://doi.org/10.1038/nbt0698-566>
  35. Li G-M. 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res* 18:85–98. <https://doi.org/10.1038/cr.2007.115>
  36. Lu AL, Clark S, Modrich P. 1983. Methyl-directed repair of DNA base-pair mismatches *in vitro*. *Proc Natl Acad Sci U S A* 80:4639–4643. <https://doi.org/10.1073/pnas.80.15.4639>
  37. Lee H, Popodi E, Tang H, Foster PL. 2012a. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* 109:E2774–83. <https://doi.org/10.1073/pnas.1210309109>
  38. Lee H, Popodi E, Tang H, Foster PL. 2012b. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* 109:E2774–83. <https://doi.org/10.1073/pnas.1210309109>
  39. Posnick LM, Samson LD. 1999. Influence of S-Adenosylmethionine pool size on spontaneous mutation, *dam* methylation, and cell growth of *Escherichia coli*. *J Bacteriol* 181:6756–6762. <https://doi.org/10.1128/JB.181.21.6756-6762.1999>
  40. Westphal LL, Sauvey P, Champion MM, Ehrenreich IM, Finkel SE. 2016. Genomewide *dam* methylation in *Escherichia coli* during long-term stationary phase. *mSystems* 1:e00130-16. <https://doi.org/10.1128/mSystems.00130-16>
  41. Walworth NG, Lee MD, Dolzhenko E, Fu F-X, Smith AD, Webb EA, Hutchins DA. 2021. Long-term m5C methylome dynamics parallel phenotypic adaptation in the cyanobacterium *Trichodesmium*. *Mol Biol Evol* 38:927–939. <https://doi.org/10.1093/molbev/msaa256>
  42. Behringer MG, Choi BI, Miller SF, Doak TG, Karty JA, Guo W, Lynch M. 2018. *Escherichia coli* cultures maintain stable subpopulation structure during long-term evolution. *Proc Natl Acad Sci U S A* 115:E4642–E4650. <https://doi.org/10.1073/pnas.1708371115>
  43. Stoiber M. 2019. Megalodon (0.1.0). GitHub. Available from: <https://github.com/nanoporetech/megalodon>
  44. Yuen Z-S, Srivastava A, Daniel R, McNevin D, Jack C, Eyras E. 2021. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat Commun* 12:3438. <https://doi.org/10.1038/s41467-021-23778-6>
  45. Liu Q, Fang L, Yu G, Wang D, Xiao C-L, Wang K. 2019. Detection of DNA base modifications by deep recurrent neural network on Oxford nanopore sequencing data. *Nat Commun* 10:2449. <https://doi.org/10.1038/s41467-019-10168-2>
  46. Ni P, Huang N, Zhang Z, Wang D-P, Liang F, Miao Y, Xiao C-L, Luo F, Wang J. 2019. DeepSignal: detecting DNA methylation state from nanopore sequencing reads using deep-learning. *Bioinformatics* 35:4586–4595. <https://doi.org/10.1093/bioinformatics/btz276>
  47. Stoiber M, Quick J, Egan R, Eun Lee J, Celniker S, Neely RK, Loman N, Pennacchio LA, Brown J. 2017. *De novo* Identification of DNA modifications enabled by genome-guided nanopore signal processing. *Bioinformatics*. <https://doi.org/10.1101/094672>
  48. Wang MX, Church GM. 1992. A whole genome approach to *in vivo* DNA-protein interactions in *E. coli*. *Nature* 360:606–610. <https://doi.org/10.1038/360606a0>
  49. Van der Woude M, Hale WB, Low DA. 1998. Formation of DNA methylation patterns: nonmethylated GATC sequences in *gut* and *pap* Operons. *J Bacteriol* 180:5913–5920. <https://doi.org/10.1128/JB.180.22.5913-5920.1998>
  50. Tierrafría VH, Rioualen C, Salgado H, Lara P, Gama-Castro S, Lally P, Gómez-Romero L, Peña-Loredo P, López-Almazo AG, Alarcón-Carranza G, Betancourt-Figueroa F, Alquicira-Hernández S, Polanco-Morelos JE, García-Sotelo J, Gaytan-Nuñez E, Méndez-Cruz C-F, Muñiz LJ, Bonavides-Martínez C, Moreno-Hagelsieb G, Galagan JE, Wade JT, Collado-Vides J. 2022. Regulondb 11.0: comprehensive high-throughput datasets on transcriptional regulation in *Escherichia Coli* K-12. *Microb Genom* 8:mgen000833. <https://doi.org/10.1099/mgen.0.000833>
  51. Cohen NR, Ross CA, Jain S, Shapiro RS, Gutierrez A, Belenky P, Li H, Collins JJ. 2016. A role for the bacterial GATC methylome in antibiotic stress survival. *Nat Genet* 48:581–586. <https://doi.org/10.1038/ng.3530>
  52. Sánchez-Romero MA, Olivenza DR, Gutiérrez G, Casadesús J. 2020. Contribution of DNA adenine methylation to gene expression heterogeneity in *Salmonella enterica*. *Nucleic Acids Res* 48:11857–11867. <https://doi.org/10.1093/nar/gkaa730>
  53. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* 1:147. <https://doi.org/10.1038/ncomms1146>
  54. Ramisetty BCM, Sudhakari PA. 2019. “Bacterial ‘grounded’ prophages: hotspots for genetic renovation and innovation”. *Front Genet* 10:65. <https://doi.org/10.3389/fgene.2019.00065>
  55. Siguier P, Goureyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 38:865–891. <https://doi.org/10.1111/1574-6976.12067>
  56. Lee H, Doak TG, Popodi E, Foster PL, Tang H. 2016. Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. *Nucleic Acids Res* 44:7109–7119. <https://doi.org/10.1093/nar/gkw647>
  57. Naas T, Blot M, Fitch WM, Arber W. 1994. Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. *Genetics* 136:721–730. <https://doi.org/10.1093/genetics/136.3.721>
  58. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13:R87. <https://doi.org/10.1186/gb-2012-13-10-r87>
  59. Qiao J, Tan X, Ren H, Wu Z, Hu X, Wang X, Atomi H. 2021. Construction of an *Escherichia coli* strain lacking Fimbriae by deleting 64 genes and its

- application for efficient production of Poly(3-Hydroxybutyrate) and L-Threonine. *Appl Environ Microbiol* 87. <https://doi.org/10.1128/AEM.00381-21>
60. Henderson IR, Meehan M, Owen P. 1997. A novel regulatory mechanism for a novel phase-variable outer membrane protein of *Escherichia coli*. *Adv Exp Med Biol* 412:349–355. [https://doi.org/10.1007/978-1-4899-1828-4\\_56](https://doi.org/10.1007/978-1-4899-1828-4_56)
  61. Graveline R, Mourez M, Hancock MA, Martin C, Boisclair S, Harel J. 2011. Lrp–DNA complex stability determines the level of ON cells in type P fimbriae phase variation. *Mol Microbiol* 81:1286–1299. <https://doi.org/10.1111/j.1365-2958.2011.07761.x>
  62. Wallecha A, Munster V, Correnti J, Chan T, van der Woude M. 2002. Dam- and OxyR-Dependent phase variation of *agn43*: essential elements and evidence for a new role of DNA methylation. *J Bacteriol* 184:3338–3347. <https://doi.org/10.1128/JB.184.12.3338-3347.2002>
  63. Stevenson G, Neal B, Liu D, Hobbs M, Packer NH, Batley M, Redmond JW, Lindquist L, Reeves P. 1994. Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J Bacteriol* 176:4144–4156. <https://doi.org/10.1128/jb.176.13.4144-4156.1994>
  64. Eguchi Y, Utsumi R. 2014. Alkali metals in addition to acidic pH activate the EvgS Histidine kinase sensor in *Escherichia coli*. *J Bacteriol* 196:3140–3149. <https://doi.org/10.1128/JB.01742-14>
  65. Kern R, Malki A, Abdallah J, Tagourt J, Richarme G. 2007. *Escherichia coli* HdeB is an acid stress chaperone. *J Bacteriol* 189:603–610. <https://doi.org/10.1128/JB.01522-06>
  66. Yasui R, Washizaki A, Furihata Y, Yonesaki T, Otsuka Y. 2014. AbpA and AbpB provide anti-phage activity in *Escherichia coli*. *Genes Genet Syst* 89:51–60. <https://doi.org/10.1266/ggs.89.51>
  67. Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X. 2006. Structure and substrate recognition of the *Escherichia coli* DNA adenine methyltransferase. *J Mol Biol* 358:559–570. <https://doi.org/10.1016/j.jmb.2006.02.028>
  68. Maier JAH, Jeltsch A. 2018. Design and Application of 6mA-Specific Zinc-Finger Proteins for the Readout of DNA Methylation, p. 29–41. In Liu, J (ed.), *Zinc finger proteins: methods and protocols*. Springer, New York, NY.
  69. Sera T, Uranga C. 2002. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry* 41:7074–7081. <https://doi.org/10.1021/bi020095c>
  70. Nabel CS, Manning SA, Kohli RM. 2012. The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential. *ACS Chem Biol* 7:20–30. <https://doi.org/10.1021/cb2002895>
  71. Shapiro R, Klein RS. 1966. The deamination of cytidine and cytosine by acidic buffer solutions. *Biochemistry* 5:2358–2362. <https://doi.org/10.1021/bi00871a026>
  72. Lindahl T, Nyberg B. 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* 13:3405–3410. <https://doi.org/10.1021/bi00713a035>
  73. O’Brown ZK, Greer EL. 2016. N6-Methyladenine: A Conserved and Dynamic DNA Mark, p. 213–246. In Jeltsch, A, Jurkowska, RZ (eds.), *DNA Methyltransferases – Role and Function*. Springer International Publishing, Cham.
  74. Gardner RS, Wahba AJ, Basilio C, Miller RS, Lengyel P, Speyer JF. 1962. Synthetic polynucleotides and the amino acid code, V. *Proc Natl Acad Sci U S A* 48:2087–2094. <https://doi.org/10.1073/pnas.48.12.2087>
  75. Sater MRA, Lamelas A, Wang G, Clark TA, Röltgen K, Mane S, Korlach J, Pluschke G, Schmid CD. 2015. DNA methylation assessed by SMRT sequencing is linked to mutations in *Neisseria meningitidis* isolates. *PLoS ONE* 10:e0144612. <https://doi.org/10.1371/journal.pone.0144612>
  76. Brunet YR, Bernard CS, Cascales E. 2020. Fur-dam regulatory interplay at an internal promoter of the Enteroaggregative *Escherichia coli* type VI secretion *Sci1* gene cluster. *J Bacteriol* 202:e00075-20. <https://doi.org/10.1128/JB.00075-20>
  77. Mehling JS, Lavender H, Clegg S. 2007. A dam methylation mutant of *Klebsiella pneumoniae* is partially attenuated. *FEMS Microbiol Lett* 268:187–193. <https://doi.org/10.1111/j.1574-6968.2006.00581.x>
  78. McIntyre ABR, Alexander N, Grigorev K, Bezdan D, Sichtig H, Chiu CY, Mason CE. 2019. Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. *Nat Commun* 10:579. <https://doi.org/10.1038/s41467-019-08289-9>
  79. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms SnpEff. *Fly* 6:80–92. <https://doi.org/10.4161/fly.19695>
  80. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
  81. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeda D, García-Sotelo JS, Alquicira-Hernández K, Muñoz-Rascado LJ, Peña-Loredo P, Ishida-Gutiérrez C, Velázquez-Ramírez DA, Del Moral-Chávez V, Bonavides-Martínez C, Méndez-Cruz C-F, Galagan J, Collado-Vides J. 2019. RegulonDB V 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res* 47:D212–D220. <https://doi.org/10.1093/nar/gky1077>
  82. Wang H-Q, Tuominen LK, Tsai C-J. 2011. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* 27:225–231. <https://doi.org/10.1093/bioinformatics/btq650>
  83. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, Lin SM. 2010. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11:587. <https://doi.org/10.1186/1471-2105-11-587>
  84. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Stat Methodol* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
  85. Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185–193. <https://doi.org/10.1093/bioinformatics/19.2.185>
  86. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’Hara RB, Solymos P, Stevens MHH, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, Caceres MD, Durand S, Evangelista HBA, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill MO, Lahti L, McGlenn D, Ouellette M-H, Cunha ER, Smith T, Stier A, Braak C, Weedon J. 2022. *Vegan: community ecology package*. GitHub. Available from: <https://github.com/vegandevs/vegan>
  87. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2:100141. <https://doi.org/10.1016/j.xinn.2021.100141>
  88. Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16:284–287. <https://doi.org/10.1089/omi.2011.0118>
  89. Carlson M. 2019. *Org.Eck12.eg.db: genome wide annotation for E. coli strain K12 (3.8.2)*.