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

Authors: Carl J. Stone[1], Gwyneth F. Boyer[2], and Megan G Behringer[1,3*]

Affiliations:

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA; Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA

[2]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, AZ

[3]Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232, USA; Vanderbilt Institute of Infection, Immunology, and Inflammation, Vanderbilt University Medical Center, Nashville, TN 37232, USA; Vanderbilt Microbiome Innovation Center, Vanderbilt University, Nashville, TN 37235, USA

*Corresponding Author:

Megan G. Behringer

Vanderbilt University

Department of Biological Sciences

VU Station B, Box 35-1634

Nashville, TN 37235-1634

Previous version available on BioRxiv: https://doi.org/10.1101/2022.12.14.520158

Classification: Biological Sciences - Evolution

Keywords: experimental evolution, epigenetics, *Escherichia coli*, N6-methyladenine, mutation

ABSTRACT

Methylated DNA adenines (6mA) are an important epigenetic modification in bacteria that affect varied 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. Utilizing third-generation sequencing technology, we produced a detailed analysis of 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-knockout (MMR-) populations after 2400 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.

MAIN TEXT

INTRODUCTION

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 6mA, 5mC, and N^4-methylcytosine (4mC), which are added by DNA methyltransferases either as part of a restriction-modification system or by solitary methyltransferases without associated 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[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]. 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[18,22,38].

53 Given the role of MMR in preventing replicative errors from becoming permanent mutations, we
54 hypothesize that maintaining MMR imposes selective pressure on the *E. coli* genome to maintain GATC
55 sites and associated 6mA modifications. While 6mA can have local gene expression effects when GATC
56 sites coincide with regulatory elements and transcription factor binding sites, MMR is likely the most
57 significant use of (and influence on) 6mA genome-wide. Thus, we predict that in the absence of MMR,
58 *E. coli* experimentally evolved for thousands of generations will exhibit increased variability in per-site
59 6mA methylation and a genome-wide decrease in 6mA due to a sustained accumulation of epimutations
60 that do not incur the strongly deleterious fitness effect of causing MMR dysfunction. There has been
61 limited examination of the dynamics of 6mA in evolving bacterial populations[39,40]. Thus, in this study, we
62 sought to investigate how the presence and absence of MMR influence 6mA methylation in evolving
63 bacterial populations over an extended time period. To accomplish this, we created a differential
64 methylation-calling pipeline designed for bacterial 6mA data from third-generation sequencing: CoMMA
65 (Comparison of Microbial Methylated Adenines). Third-generation sequencing platforms (such as
66 PacBio's SMRT sequencing or ONT's Nanopore sequencing) allow for accurate and cost-effective
67 identification of DNA base modifications. Using Nanopore sequencing and CoMMA, we first characterized
68 genome-wide methylation frequency at GATC sites in *E. coli* MG1655. Then, focusing on clones isolated
69 from evolved populations with initially WT and MMR-deficient genetic backgrounds, we characterized
70 how 6mA methylation frequency changed among all 19,120 GATC motifs after 2000 generations of
71 experimental evolution. We then compared our findings to studies of spontaneous mutation rates in *E.*
72 *coli* to determine how MMR constrains genetic and epigenetic evolution at GATC sites. Together, our
73 results illustrate the evolvability of 6mA and its potential to contribute to heritable change in bacterial
74 populations.
75
76 **RESULTS**
77 **6mA methylation status varies between genomic features in *E. coli* MG1655.**
78 We first measured genome-wide methylation in *E. coli* K-12 MG1655 (strain PMF2) to
79 characterize methylation trends in our wild-type ancestor. Briefly, three colonies from an overnight culture

80 were sequenced on an ONT MinION portable sequencer. Nanopore sequencing produced an average
81 read depth of ~50 reads per GATC site, and there was no significant strand bias in coverage (**Fig. S2**).
82 To measure methylation from Nanopore reads, we used the methylation classification program
83 Megalodon to calculate the proportion of methylated reads to total reads at each GATC site, called the
84 percent methylation (**Dataset S1**)[41]. Previous benchmarking studies have shown Megalodon to be one
85 of the most reliable methylation callers when applied to data generated from Nanopore sequencing[41–43].
86 To confirm this we compared the performance of multiple independently developed methylation calling
87 programs, and with our data Megalodon consistently identified most of the same sites as methylated or
88 unmethylated and correctly identified stably unmethylated sites reported previously (see **Methods,**
89 **Supplementary Text 1, Fig. S3, Fig. S4, Table S1, Table S2**)[34,44–47]. After classifying methylated sites
90 with Megalodon, all GATC sites were annotated with data from RegulonDB, a curated database of *E. coli*
91 K-12 transcriptional regulatory networks[48]. Consistent with previous reports, most GATC sites were
92 almost completely methylated, with a median genome-wide percent methylation of 97% (**Fig. 1A**)[8,34,39,49].
93 Methylation profiling revealed that 177 double-stranded GATC sites are hypomethylated on at least one
94 strand, of which 138 were hemimethylated and 39 were hypomethylated on both strands (**Table S3**).
95 Intergenic GATC sites were significantly less methylated compared to those within coding sequences
96 (median percent methylated reads of 93% compared to 97%; Wilcoxon rank sum test, $P < 2.2\times10^{-16}$).
97 This decrease at intergenic sites is likely due to the higher proportion of transcriptional regulatory regions
98 and transcription factor binding sites in intergenic regions, where these sites may exclude binding of the
99 Dam methyltransferase to GATC sites.

100 For a broader view of the factors that influence GATC methylation and to determine if
101 hypomethylation was associated with GATC sites that may be shielded by protein binding, we looked at
102 sites within intergenic regions where DNA-protein interactions likely occur, such as sigma factor binding
103 sites and binding sites of other transcriptional regulatory proteins[2,33,46,47,50]. Overall, within these
104 intergenic regions, there is a general decrease in methylation around promoters. GATC sites within -10
105 and -35 promoter elements were significantly less methylated (95.2% for -10 elements, $P_{-10} = 6.7\times10^{-8}$;
106 90.4% for -35 elements, $P_{-35} = 6.7\times10^{-10}$, Wilcoxon signed rank test with Benjamini-Hochberg correction)

107 than the genome-wide median percent methylation (96.6%). This is best observed when methylation is
108 averaged across all promoter regions in the genome; there is a lower frequency of methylated reads
109 around 100 bp before and after the transcription start site, with a maximum decrease in methylation of
110 ~5% centered on the -35 element (**Fig 2A**). Moreover, -35 promoter elements known to be bound by σ^{70},
111 σ^{38}, and σ^{24} had significantly lower percent methylation than the genome-wide median (Wilcoxon signed
112 rank test with BH correction: $P_{sigma70} = 2.0\times10^{-10}$; $P_{sigma38} = 7.0\times10^{-4}$; $P_{sigma24} = 1.6\times10^{-5}$) (**Fig 2B**).
113 In addition to core promoter elements, binding sites of other transcriptional regulatory proteins
114 within intergenic regions also exhibited differences in their degree of hypomethylation (**Fig 2C**).
115 Specifically, the global regulators CRP, Fis, Fnr, and Nac were significantly less methylated than the
116 genome-wide median (Wilcoxon signed rank test with BH correction: $P_{CRP} = 2.0\times10^{-18}$; $P_{Fis} = 4.2\times10^{-3}$;
117 $P_{Fnr} = 9.1\times10^{-6}$; $P_{Nac} = 4.98\times10^{-2}$). Differences in methylation were also observed within coding regions
118 for a few genomic features (**Fig 2C**). Across the 12 cryptic prophages in the *E. coli* K-12 genome,
119 significantly decreased methylation is observed in 9 of them: CP4-44, CP4-57, CP4-6, CPS-53, DLP12,
120 e14, Qin, Rac, and PR-Y (CP4-44: 94.0% methylation, $P = 6.2\times10^{-4}$; CP4-57: 92.3%, $P = 9.4\times10^{-16}$; CP4-
121 6: 95.3%, $P = 2.5\times10^{-13}$; CPS-53: 92.4%, $P = 4.7\times10^{-6}$; DLP12: 88.5%, $P = 3.7\times10^{-13}$; e14: 92.1%, P =
122 1.1×10^{-8}; Qin: 95.1%, $P = 2.9\times10^{-6}$; Rac: 94.9%, $P = 4.3\times10^{-8}$; PR-Y: 93.8%, $P = 2.6\times10^{-2}$; Wilcoxon
123 signed rank test with BH correction)[51,52]. Of the various IS elements, *IS*150, *IS*4, *IS*30A, *IS*30C, and *IS*30D
124 had significantly decreased methylation (*IS*150: 89.4% methylation, $P = 2.6\times10^{-3}$; *IS*4: 94.1%, $P = 1.3\times10^{-}$
125 ^{2}; *IS*30A: 89.7%, $P = 4.3\times10^{-2}$; *IS*30C: 85.8%, $P = 1.5\times10^{-2}$; *IS*30D: 86.7%, $P = 2.2\times10^{-2}$)[53–55]. Lastly,
126 some genomic features, such as REP elements and terminator sequences, do not show measurable
127 differences in 6mA methylation compared to the genome-wide average. However, these regions either
128 do not contain many GATC sites or are poorly annotated. Comparing sequencing replicates from future
129 studies across a variety of conditions, combined with improving genomic annotations, could allow for the
130 detection of higher-resolution methylation differences and the characterization of methylation on
131 abundant but unannotated sites.
132
133

134 **Differential methylation analysis shows that MMR- clones evolve decreased GATC methylation.**

135 After characterizing a baseline *E. coli* K-12 substr. MG1655 6mA methylome, we compared
136 GATC methylation across the experimental ancestor PMF2 strain and eight isolated clones from
137 populations that were experimentally evolved for 2400 generations with either intact (WT) or disrupted
138 mismatch repair (MMR-)[56]. Briefly, CoMMA adapts the differential methylation package methylKit to
139 identify GATC sites with significant differences in methylation status between bacterial samples using
140 logistic regression[57]. For a broad view of how methylation evolves, we focused on two populations from
141 each of the WT or MMR- genetic backgrounds; from each of these populations, we selected two clones
142 and ensured each clone belonged to a different evolved subpopulation. This sampling scheme would
143 allow us to evaluate how methylation evolves within and among populations depending on their ability to
144 conduct methyl-directed mismatch repair.

145 As with the ancestor, experimentally evolved clones were Nanopore sequenced, and methylated
146 sites were called using Megalodon (**Dataset S2**). The median genome-wide methylation percentage of
147 both genetic backgrounds decreased during experimental evolution with MMR- lines exhibiting a
148 significantly larger median decrease in methylation (-0.15%) compared to evolved WT lines (-0.03%;
149 Wilcoxon rank sum test, $P < 2.2\times10^{-16}$, **Fig. S5**). Next, we determined which specific GATC sites were
150 differentially methylated between the evolved lines and the ancestor (**Fig 3A-B**). Using cutoffs $P < 0.05$
151 and percent methylation change >10%, there were significantly more differentially methylated sites in the
152 MMR- lines (28 sites with increased methylation, 141 sites with decreased methylation) compared to the
153 evolved WT lines (22 increased, 16 decreased; Fisher's exact test, $P < 2.2\times10^{-16}$).

154 Both evolved backgrounds also had a significant overrepresentation of differentially methylated
155 sites within certain genomic regions, especially in sites where DNA-protein interactions occur. Across the
156 genome, differentially methylated sites in both WT and MMR- evolved clones appear to coincide with
157 regions of low methylation in the ancestor and with cryptic prophages and IS elements (**Fig. 1B-C**). Sites
158 that were hemimethylated and hypomethylated in the MG1655 ancestor were enriched for differential
159 methylation in the evolved clones (Fisher's exact test, $P < 2.2\times10^{-16}$; **Table S3**). Differentially methylated
160 sites in both evolved backgrounds were significantly over-enriched in transcription factor binding motifs

161 and IS elements (Fisher's exact test with FDR correction, $P < 0.05$, **Table 1**). Sites with increased
162 methylation in both evolved backgrounds were also over-enriched within cryptic prophages, and sites
163 with decreased methylation were over-enriched in Rho-independent terminators. The MMR- evolved lines
164 also showed enrichment for under-methylated sites within cryptic prophages ($P < 2.2\times10^{-16}$) and within
165 promoters, specifically σ^{70}-binding sites in the -35 region ($P < 2.2\times10^{-16}$).

166 Gene ontology (GO) enrichment analysis of genes containing or near differentially methylated
167 sites revealed enrichment in genes associated with stress response, biofilm formation, and virulence
168 factor production. Of the 194 differentially methylated sites identified between both WT and MMR-
169 evolved clones, 13 sites were in the same gene or intergenic region between the two genotypes (**Fig.**
170 **4A**). Ten of these sites were in the same palindromic GATC site, and 7 of those were in the same adenine
171 residue. GO enrichment analysis was then done on genes that contained differentially methylated GATC
172 sites, or if the GATC site was not within a gene then the nearest gene was used. Genes that were less
173 methylated in the MMR- clones were significantly enriched for biofilm formation and transcriptional
174 regulators (**Fig. 4B**). Genes that were more methylated after evolution were enriched for biofilm formation
175 in the evolved WT clones and for lipopolysaccharide metabolism in both WT and MMR- clones.
176 Interestingly, six of the fifteen biofilm genes identified belong to three different putative chaperone-usher
177 fimbrial operons. These fimbriae are cryptic in *E. coli* K-12 under laboratory conditions, but when
178 overexpressed they increase biofilm formation[58]. Other biofilm genes *yfaL* and *ycgV* are homologs of the
179 cryptic prophage-encoded autotransporter and adhesin Antigen 43 (Ag43 encoded by *flu*)[59]. Both fimbriae
180 and Ag43 are known to be subject to methylation-regulated phase variation in *E. coli* and related
181 species[31,59–61]. Other enriched gene ontologies include lipopolysaccharide synthesis (including the non-
182 functional O-antigen polymerase *wbbH*), acid stress response genes, phage defense genes, and various
183 transcriptional regulators involved in stress response (**Table S4**)[62–65].

184 Comparison of the full 6mA methylomes between evolved clones and their ancestor revealed a
185 non-deterministic change in 6mA in the absence of MMR consistent with random genetic drift. The
186 percent of methylated reads at all GATC sites was compared between samples by clustering using non-
187 metric multidimensional scaling (NMDS; **Fig. 3C**). The evolved WT clones and their ancestor clustered

188 together, showing that their 6mA methylomes remain similar after extended evolution (Pairwise
189 PERMANOVA Ancestor vs. WT, P = 0.168; Analysis of multivariate homogeneity of group dispersions,
190 ANOVA with Tukey's HSD: $P_{Ancestor\text{-}WT}$ = 0.99). In contrast, MMR- evolved clones diverge from the
191 ancestor in random directions, and the MMR- clones do not cluster together. The centroid of the MMR-
192 clones was near that of the ancestor and WT clones but significantly different than the WT clones, and
193 MMR- clone dispersions were significantly greater than both the ancestor and WT groups (Pairwise
194 PERMANOVA WT vs. MMR-, P = 0.026; PERMDISP2, ANOVA with Tukey's HSD: $P_{Ancestor\text{-}MMR-}$ = 0.018,
195 $P_{WT\text{-}MMR-}$ = 0.010). This suggests that 6mA differences that occur during evolution in MMR- lines are non-
196 deterministic, resulting in significant changes to methylation for each sample but not strongly in any one
197 direction for the whole group. MMR- clones from the same population did not cluster together, indicating
198 that shared evolutionary history does not have a large effect on broad 6mA differences over this time
199 scale. Together, the significant number of under-methylated sites and the random divergence of
200 differentially methylated sites in the MMR- clones compared to the ancestor suggest that 6mA is shaped
201 by genetic drift in the absence of MMR.
202
203 **MMR- evolved lines accumulate GATC mutations at a rate that indicates relaxed selection.**
204 After identifying decreased maintenance of 6mA in the absence of MMR, we sought to
205 genomically identify whether GATC sites in MMR- clones were also under decreased evolutionary
206 constraint. Here, we characterized the fraction of mutations that altered GATC sites across the
207 experimentally evolved clones from both evolved genetic backgrounds. After ~2400 generations, MMR-
208 clones had acquired a total of 959 mutations with 34 mutations (3.5% of mutations) causing a gain or
209 loss of a GATC site (**Table 2, Table S4**). In contrast, the WT clones accumulated a total of 33 mutations,
210 none of which had affected GATC sites. Since the relative lack of mutations in WT clones limited our
211 statistical power, we compared the mutation rates of the experimentally evolved clones to mutation rates
212 calculated in a minimally-selective mutation accumulation (MA) study by Lee *et al.*[37]. As this study directly
213 measured the mutation rates of the same WT and MMR- PFM2 strain, we are able to accurately
214 determine the expected rate of GATC mutation in the absence of selection[37].

215 Reanalysis of the MA data to calculate mutation rates in tetranucleotide contexts determined that
216 3.5% of mutations that arose in MMR- lines during MA were in GATC sites, and is equal to the fraction
217 of mutations affecting GATC sites in the MMR- lines during experimental evolution[37]. The similarity of
218 these GATC mutation rates suggests that experimentally evolved MMR- clones acquire mutations in
219 GATC sites at the neutral rate and may be evolving under relaxed selection. Additional evidence that
220 GATC sites and associated adenine methylation may be evolving under relaxed selection in
221 experimentally evolved MMR- clones include that the clones isolated from the MMR- Population 113 both
222 acquired mutations resulting in a nonsynonymous D20A substitution proximal to the GATC recognition
223 domain of *dam* methyltransferase[66]. While the exact consequences of this mutation have not been
224 examined, this substitution could affect DNA binding and GATC recognition, indicating that perfect Dam
225 function may not be necessary in the absence of MMR. Together, these data suggest that MMR is a
226 major factor in maintaining GATC sites and *dam* methyltransferase, and in the absence of MMR these
227 loci are vulnerable to mutation.
228
229 **DISCUSSION**
230 In this study, we present what is to our knowledge a first-of-its-kind high-resolution
231 characterization of the *E. coli* K-12 MG1655 6mA methylome that includes a functional and genic
232 classification of 6mA methylation. We identified patterns of reduced methylation at intergenic regions,
233 promoter elements, DNA binding sites, and cryptic prophages. This reduced methylation could result
234 from the physical exclusion of Dam methyltransferase at these sites and potentially cause downstream
235 effects on gene regulation and stress response. The ability to acquire high-resolution methylation data
236 opens the door for future investigations on the specific effects of these methylation differences when
237 paired with transcriptomic and biochemical experiments.
238 We also created and applied the CoMMA pipeline to investigate the effect of 2400 generations of
239 evolution on WT and MMR- *E. coli* MG1655 clones. We found that the MMR- clones exhibited a significant
240 decrease in methylation, indicating that 6mA is somewhat dispensable in the absence of MMR. While
241 methylation was largely stable at the majority of GATC sites, MMR- clones exhibited significantly more

242 differentially methylated sites than WT clones. Interestingly, these differentially methylated sites are often
243 located in DNA binding sites. Adenine modification can affect specific binding affinity where 6mA
244 coincides with a DNA binding site[67,68]. Without the genome-wide pressure to maintain 6mA to minimize
245 mutation rates, MMR- clones are free to alter 6mA at these sites with a bias toward decreased
246 methylation, likely due to binding competition between transcription factors and the methyltransferase.
247 Altered 6mA methylation that affects transcription factor binding can result in downstream effects and
248 further alter cellular phenotypes. Our study demonstrates that differential methylomics is a powerful
249 approach that can directly resolve differences between clones that are not apparent from genetic changes
250 alone. Thus, the CoMMA pipeline is a useful tool that can complement genetic approaches.

251 An interesting observation from our study involves one of the MMR- evolved populations (P113),
252 which acquired a mutation in the *dam* adenine methyltransferase resulting in a single amino acid
253 substitution (DamD20A). The consequence of this mutation is unclear, although this residue is proximal
254 to three S-adenosylmethionine-binding sites at K10, K14, and D54, and could potentially affect protein
255 folding[66]. While there are no broad methylation differences within these clones compared to others of the
256 same background, a DamD20A substitution may have contributed to the observed changes in
257 methylation in these clones. Further studies are needed to investigate if this mutation altered Dam activity
258 and how this affects genome-wide methylation patterns.

259 One potential impact of our findings is how 6mA methylation may influence the mutational
260 spectrum and adaptive landscape of bacteria. DNA base modifications can affect the rates of
261 spontaneous deamination; for instance, 5mC is vulnerable to deamination and contributes to the C -> T
262 mutation rate[69–71]. Similarly, adenine and 6mA may deaminate to hypoxanthine, which pairs with cytosine
263 during replication[72,73]. While the spontaneous deamination rates of adenine versus 6mA have not been
264 exhaustively studied, recent work identified a positive correlation between 6mA and A→G transitions in
265 *Neisseria meningitidis*[74]. Thus, decreases in 6mA after MMR loss may result in a decreased A→G
266 transition rate. This change to the mutational spectrum could potentially affect how MMR- or Dam- cells
267 adapt to changing environments and may influence the trajectory of their evolutionary paths.

268 This study is the first-ever analysis of differential methylation between laboratory evolved bacterial
269 lineages at this time scale. The ability to identify changes to 6mA at GATC sites over evolutionary time
270 creates opportunities for the seamless integration of genetic and epigenetic data. Nanopore sequencing
271 and pipelines like CoMMA are sufficiently reliable, affordable, and high-throughput to powerfully analyze
272 genetic and epigenetic evolution in bacteria. As more studies uncover how bacterial DNA methylation
273 affects gene expression, differential epigenetic analyses can be extended through time or between
274 bacterial isolates and help to elucidate the relationship between mutation, epimutation, and gene
275 expression as it relates to adaptation to challenging environments.[39,40,75,76].
276
277 **METHODS**
278 **Bacterial strains and culture conditions**
279 All bacterial strains were cultured and sequenced to identify mutations as part of a long-term
280 experimental evolution study that has been described previously[56]. Briefly, the WT ancestor strain PFM2
281 is a prototrophic derivative of *E. coli* K-12 MG1655, and the MMR-deficient strain PFM5 is derived from
282 PFM2 with an in-frame deletion of the *mutL* gene[56]. Long-term cultures were started by inoculating 10
283 mL of LB-Miller broth (BD Difco) in 16- x 100-mm glass culture tubes with a single colony of the progenitor
284 strain[56]. Cultures were maintained shaking at 175 rpm, alternating transfers after 24 h of growth at 37 °C
285 and 48 h of growth at 25 °C. Cultures of interest in this study were transferred by thoroughly vortexing
286 the culture tube and inoculating 9 mL of LB-Miller broth with 1 mL of culture, resulting in a transfer of
287 about 10^9 cells[56]. Populations were grown and transferred for 2000 total generations of growth. After
288 2000 generations, samples from the evolving populations were streaked for isolation and eight random
289 clones were saved in 40% glycerol at -80 °C for further and future experimentation.
290
291 **Nucleic acid isolation and Nanopore sequencing**
292 Evolved and WT ancestor clones were revived from frozen storage on LB agar by incubating
293 overnight at 37 °C. For each clone, a single isolated colony was randomly selected and restreaked on to
294 fresh LB agar and again incubated overnight at 37 °C for DNA extraction. Following overnight incubation,

295 colonies were collected from the agar using a 10mL loop and inoculated directly into DNA extraction
296 buffer and DNA was further isolated with the UltraClean Microbial DNA Kit (Qiagen). Following
297 recommendations from ONT, Extracted DNA was assessed for quantity using the Qubit High Sensitivity
298 dsDNA kit and quality using the Genomic DNA ScreenTape on a TapeStation 4200 (Agilent). Libraries
299 were then prepared for Nanopore sequencing using the PCR-free Rapid DNA sequencing Kit (SQK-
300 RAD004, Nanopore) paired with the Rapid Barcoding Kit (SQK-RBK004). Prepared libraries were pooled
301 and loaded on to a R.9.4.1 flow cell and sequenced on a MinION for 48h with live basecalling in
302 MinKNOW deactivated.
303
304 **Bioinformatics and sequencing analysis**
305 To determine the precision of methylation calling, we compared results for each of four
306 methylation calling pipelines across three Nanopore sequencing replicates of the WT ancestor PFM2 and
307 determined the number of inconsistently called sites. Initial Nanopore basecalling by converting fast5 files
308 into fastq was conducted using GuppyGPU (v. 3.1.5+781ed57). Following basecalling, samples were
309 demultiplexed with Guppy GPU. Average read length across all samples is 5583 +/- 275 nucleotides (+/-
310 s.e.m). Following basecalling and demultiplexing, we evaluated four different pipelines (Tombo (v.1.5),
311 mCaller (v.0.3), Deepsignal (v.0.1.6), and Megalodon (v.0.1.0)) for their precision and accuracy to identify
312 base modifications and determine methylated adenines within GATC sites[41,44,77,78]. Both Tombo and
313 Megalodon were created by ONT for methylation calling, with Tombo being an older release that relies
314 on statistical methods to annotate non-canonical bases and Megalodon being a newer release based on
315 machine learning with deep neural networks[41,78]. To increase accuracy, Megalodon can be run with a
316 VCF file containing the position and identity of known variants, for which we used Illumina sequencing
317 data for evolved clones that was previously published[55]. For pipelines not developed by ONT: DeepSignal
318 applies a deep neural network to data that has been preprocessed by Tombo; and mCaller applies a
319 deep neural network to data that has been preprocessed by Nanopolish. When running the mCaller
320 pipeline we used Nanopolish v. 0.11.1). Across all samples, the mean read coverage across GATC sites
321 was 57 reads per site (**Fig. S2**).

322 Identification of mutations affecting GATC sites in evolved clones was conducted using previously
323 published WGS Illumina sequencing data. To identify the rate of mutations affecting GATC sites under
324 minimal selection, we utilized previously published mutation accumulation data for the *E. coli* strains that
325 served as the experimental ancestor for the WT (PFM2) and MMR- (PMF5, PMF2:Δ*mutL*)[37]. For both the
326 experimentally evolved clones and the MA lines, mutations were called using GATK best practices and
327 the genic location of mutations was annotated using SnpEff v. 4.3 by comparing VCF files of identified
328 mutations for each clone to the *Escherichia coli* K-12 str. MG1655 reference genome (INSDC accession:
329 U00096.3)[79]. To confirm a mutation occurred in a GATC site, the tetranucleotide context of each mutation
330 was identified with the fill-fs -l 5 function from vcftools (v. 0.1.12) to annotate each variant with the flanking
331 five upstream and downstream nucleotides[80].
332 GATC sites were annotated as within different genomic features using all combined databases in
333 RegulonDB v. 10.9[81]. These annotations were concatenated into a BED file, which was used to annotate
334 GATC sites with the CoMMA function annotateMethylSites.
335
336 **Differential methylation calling**
337 Differential methylation analysis was performed in R using CoMMA, a package that serves as a
338 wrapper for methylKit v. 3.13[57]. GATC sites that had acquired mutations in any clone were removed from
339 differential methylation analysis so that only GATC sites shared between samples were considered. To
340 ensure the reliability of differential methylation analysis, CoMMA includes the following modifications to
341 the standard methylKit workflow: 1) any sites covered by fewer than 10 reads were excluded from
342 analysis; 2) methylation coverage at each site was normalized by median coverage; and 3) sites were
343 ultimately considered to be differentially methylated based on a q-value cutoff of < 0.05 and percent
344 methylation difference > 10%. Differential methylation was called separately between the ancestor and
345 either the evolved MMR- or WT clones. Differential methylation was calculated by logistic regression with
346 a sliding linear model (SLIM) to correct for multiple hypothesis testing[57,82].
347
348

349 **Statistical analysis and ordination**

350 Statistical analysis comparing percent methylation (or β-value) was performed on M-values to
351 remove the heteroscedasticity observed in β-value distributions[83]. M-values were calculated as shown in
352 Equation 1:

353 $$M = log2\left(\frac{\beta}{1-\beta+0.001}\right) \quad (1)$$

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

358 Ordination of samples was performed after normalizing β-values. Since β-values are dependent
359 on coverage, samples with low coverage stand out when visualized with non-metric ordination methods.
360 Therefore, before ordination, we normalized coverage and methylated read counts between samples to
361 reduce the effects of coverage on ordination distance. First, the coverage of all samples was normalized
362 using quantile normalization[85]. Then, a scaling equation was calculated for each sample using the simple
363 linear regression line between coverage and normalized coverage. This scaling equation was then
364 applied to each sample's coverage and methylated read count, both of which were used to calculate
365 normalized β-values. Ordination of samples was performed using the R package vegan v. 2.6-4[86]. Non-
366 metric multidimensional scaling (NMDS) was used to visualize sample clustering based on a Euclidean
367 distance matrix calculated from normalized β-values. PERMANOVA and PERMDISP2 were done with
368 vegan functions adonis2 and betadisper, respectively.

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

373

374

Data and code availability

Newly generated sequencing data for the identification of nucleotide modifications can be downloaded from SRA, BioProject PRJNA912686. Code and 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.

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 grants F32GM123703 (M.G.B.), and start-up funding from Vanderbilt University and the Vanderbilt Center for Infection, Immunology, and Inflammation.

388 **REFERENCES**

389 1. Sánchez-Romero, M. A. & Casadesús, J. The bacterial epigenome. *Nat. Rev. Microbiol.* **18**, 7–20
390 (2020).
391 2. Wion, D. & Casadesús, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions.
392 *Nat. Rev. Microbiol.* **4**, 183–192 (2006).
393 3. Seib, K. L., Srikhanta, Y. N., Atack, J. M. & Jennings, M. P. Epigenetic Regulation of Virulence and
394 Immunoevasion by Phase-Variable Restriction-Modification Systems in Bacterial Pathogens. *Annu.*
395 *Rev. Microbiol.* **74**, 655–671 (2020).
396 4. Adhikari, S. & Curtis, P. D. DNA methyltransferases and epigenetic regulation in bacteria. *FEMS*
397 *Microbiol. Rev.* **40**, 575–591 (2016).
398 5. Ardissone, S. Cell Cycle Constraints and Environmental Control of Local DNA Hypomethylation in
399 α-Proteobacteria. *PLOS Genet* (2016) doi:10.1371/journal.pgen.1006499.
400 6. Reisenauer, A., Kahng, L. S., McCollum, S. & Shapiro, L. Bacterial DNA Methylation: a Cell Cycle
401 Regulator? *J. Bacteriol.* **181**, 5135–5139 (1999).
402 7. Collier. A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl*
403 *Acad Sci USA* (2007) doi:10.1073/pnas.0708112104.
404 8. Sánchez-Romero, M. A., Cota, I. & Casadesús, J. DNA methylation in bacteria: from the methyl
405 group to the methylome. *Curr. Opin. Microbiol.* **25**, 9–16 (2015).
406 9. De Ste Croix, M. *et al.* Phase-variable methylation and epigenetic regulation by type I restriction-
407 modification systems. *FEMS Microbiol. Rev.* **41**, S3–S15 (2017).
408 10. Nye, T. M. *et al.* DNA methylation from a Type I restriction modification system influences gene
409 expression and virulence in Streptococcus pyogenes. *PLOS Pathog.* **15**, e1007841 (2019).
410 11. Oliveira, P. H. & Fang, G. Conserved DNA Methyltransferases: A Window into Fundamental
411 Mechanisms of Epigenetic Regulation in Bacteria. *Trends Microbiol.* **29**, 28–40 (2021).
412 12. Geier, G. E. & Modrich, P. Recognition sequence of the dam methylase of Escherichia coli K12 and
413 mode of cleavage of Dpn I endonuclease. *J. Biol. Chem.* **254**, 1408–1413 (1979).
414 13. Herman, G. E. & Modrich, P. Escherichia coli dam methylase. Physical and catalytic properties of

415 the homogeneous enzyme. *J. Biol. Chem.* **257**, 2605–2612 (1982).
416 14. Marinus, M. G. & Morris, N. R. Isolation of deoxyribonucleic acid methylase mutants of Escherichia
417 coli K-12. *J. Bacteriol.* **114**, 1143–1150 (1973).
418 15. Blow, M. J. *et al.* The Epigenomic Landscape of Prokaryotes. *PLOS Genet.* **12**, e1005854 (2016).
419 16. Blattner, F. R. *et al.* The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**,
420 1453–1462 (1997).
421 17. Boye, E. & Løbner-Olesen, A. The role of dam methyltransferase in the control of DNA replication
422 in E. coli. *Cell* **62**, 981–989 (1990).
423 18. Pukkila, P. J., Peterson, J., Herman, G., Modrich, P. & Meselson, M. Effects of High Levels of DNA
424 Adenine Methylation on Methyl-Directed Mismatch Repair in *Escherichia coli*. *Genetics* **104**, 571–
425 582 (1983).
426 19. Schlagman, S. L., Hattman, S. & Marinus, M. G. Direct role of the Escherichia coli Dam DNA
427 methyltransferase in methylation-directed mismatch repair. *J. Bacteriol.* **165**, 896–900 (1986).
428 20. Messer, W. & Noyer-Weidner, M. Timing and targeting: The biological functions of Dam methylation
429 in E. coli. *Cell* **54**, 735–737 (1988).
430 21. Messer, W., Bellekes, U. & Lother, H. Effect of dam methylation on the activity of the E. coli
431 replication origin, oriC. *EMBO J.* **4**, 1327–1332 (1985).
432 22. Herman, G. E. Escherichia coli K-12 clones that overproduce dam methylase are hypermutable. *J*
433 *Bacteriol* (1981) doi:10.1128/jb.145.1.644-646.1981.
434 23. Lahue, R. S. & Modrich, P. Methyl-directed DNA mismatch repair in Escherichia coli. *Mutat. Res.*
435 *Mol. Mech. Mutagen.* **198**, 37–43 (1988).
436 24. Roberts, D., Hoopes, B. C., McClure, W. R. & Kleckner, N. IS10 transposition is regulated by DNA
437 adenine methylation. *Cell* **43**, 117–130 (1985).
438 25. Tomcsanyi, T. & Berg, D. E. Transposition effect of adenine (Dam) methylation on activity of O end
439 mutants of IS50. *J. Mol. Biol.* **209**, 191–193 (1989).
440 26. Torreblanca, J. & Casadesús, J. DNA adenine methylase mutants of Salmonella typhimurium and a
441 novel dam-regulated locus. *Genetics* **144**, 15–26 (1996).

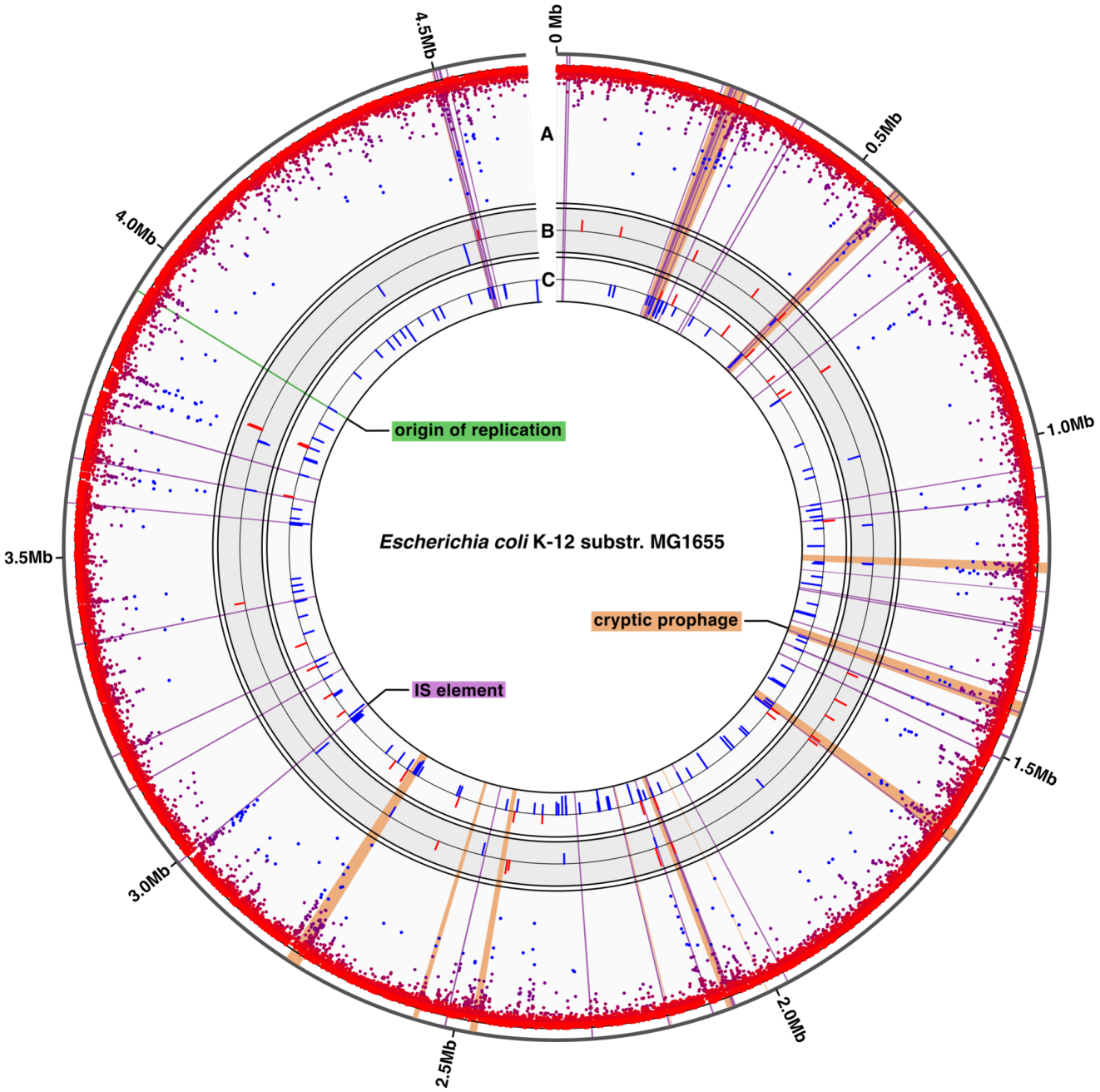
442 27. Heithoff, D. M., Sinsheimer, R. L., Low, D. A. & Mahan, M. J. An essential role for DNA adenine
443 methylation in bacterial virulence. *Science* **284**, 967–970 (1999).
444 28. Lu. SeqA: A negative modulator of replication initiation in E. coli. *Cell* (1994) doi:10.1016/0092-
445 8674(94)90156-2.
446 29. Blyn, L. B., Braaten, B. A. & Low, D. A. Regulation of pap pilin phase variation by a mechanism
447 involving differential dam methylation states. *EMBO J.* **9**, 4045–4054 (1990).
448 30. Jakomin, M., Chessa, D., Bäumler, A. J. & Casadesús, J. Regulation of the Salmonella enterica std
449 Fimbrial Operon by DNA Adenine Methylation, SeqA, and HdfR. *J. Bacteriol.* **190**, 7406–7413
450 (2008).
451 31. Hernday, A. D., Braaten, B. A. & Low, D. A. The Mechanism by which DNA Adenine Methylase and
452 PapI Activate the Pap Epigenetic Switch. *Mol. Cell* **12**, 947–957 (2003).
453 32. van der Woude, M. W. Phase variation: how to create and coordinate population diversity. *Curr.*
454 *Opin. Microbiol.* **14**, 205–211 (2011).
455 33. Sánchez-Romero, M. A. & Casadesús, J. Waddington's Landscapes in the Bacterial World. *Front.*
456 *Microbiol.* **12**, 685080 (2021).
457 34. Tavazoie, S. & Church, G. M. Quantitative whole-genome analysis of DNA-protein interactions by in
458 vivo methylase protection in E. coli. *Nat. Biotechnol.* **16**, 566–571 (1998).
459 35. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85–98 (2008).
460 36. Lu, A. L., Clark, S. & Modrich, P. Methyl-directed repair of DNA base-pair mismatches in vitro. *Proc.*
461 *Natl. Acad. Sci. U. S. A.* **80**, 4639–4643 (1983).
462 37. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous
463 mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc.*
464 *Natl. Acad. Sci.* **109**, E2774–E2783 (2012).
465 38. Posnick, L. M. & Samson, L. D. Influence of S-Adenosylmethionine Pool Size on Spontaneous
466 Mutation, Dam Methylation, and Cell Growth of Escherichia coli. *J. Bacteriol.* **181**, 6756–6762
467 (1999).
468 39. Westphal, L. L., Sauvey, P., Champion, M. M., Ehrenreich, I. M. & Finkel, S. E. Genomewide Dam

469 Methylation in Escherichia coli during Long-Term Stationary Phase. *mSystems* **1**, e00130-16
470 (2016).
471 40. Walworth, N. G. *et al.* Long-Term m5C Methylome Dynamics Parallel Phenotypic Adaptation in the
472 Cyanobacterium Trichodesmium. *Mol. Biol. Evol.* **38**, 927–939 (2021).
473 41. Oxford Nanopore Technologies. Megalodon. (2022).
474 42. Yuen, Z. W.-S. *et al.* Systematic benchmarking of tools for CpG methylation detection from
475 nanopore sequencing. *Nat. Commun.* **12**, 3438 (2021).
476 43. Liu. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore
477 sequencing data. *Nat Commun* (2019) doi:null.
478 44. Ni, P. *et al.* DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using
479 deep-learning. *Bioinformatics* **35**, 4586–4595 (2019).
480 45. Oxford Nanopore Technologies. Tombo.
481 46. Wang, M. X. & Church, G. M. A whole genome approach to in vivo DNA-protein interactions in E.
482 coli. *Nature* **360**, 606–610 (1992).
483 47. van der Woude, M., Hale, W. B. & Low, D. A. Formation of DNA Methylation Patterns:
484 Nonmethylated GATC Sequences in gut and papOperons. *J. Bacteriol.* **180**, 5913–5920 (1998).
485 48. Tierrafría, V. H. *et al.* RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional
486 regulation in Escherichia coli K-12. *Microb. Genomics* **8**, (2022).
487 49. Cohen. A role for the bacterial GATC methylome in antibiotic stress survival. *Nat Genet* (2016)
488 doi:10.1038/ng.3530.
489 50. Sanchez-Romero. Contribution of DNA adenine methylation to gene expression heterogeneity in
490 Salmonella enterica. *Nucleic Acids Res* (2020) doi:10.1093/nar/gkaa730.
491 51. Wang, X. *et al.* Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* **1**,
492 147 (2010).
493 52. Ramisetty, B. C. M. & Sudhakari, P. A. Bacterial 'Grounded' Prophages: Hotspots for Genetic
494 Renovation and Innovation. *Front. Genet.* **10**, (2019).
495 53. Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and

496 diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014).
497 54. Lee, H., Doak, T. G., Popodi, E., Foster, P. L. & Tang, H. Insertion sequence-caused large-scale
498 rearrangements in the genome of *Escherichia coli*. *Nucleic Acids Res.* gkw647 (2016)
499 doi:10.1093/nar/gkw647.
500 55. Naas, T. Insertion sequence-related genetic variation in resting Escherichia coli K-12. *Genetics*
501 (1994) doi:10.1093/genetics/136.3.721.
502 56. Behringer, M. G. *et al.* Escherichia coli cultures maintain stable subpopulation structure during long-
503 term evolution. *Proc. Natl. Acad. Sci.* **115**, E4642–E4650 (2018).
504 57. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA
505 methylation profiles. *Genome Biol.* **13**, R87 (2012).
506 58. Qiao, J. *et al.* Construction of an Escherichia coli Strain Lacking Fimbriae by Deleting 64 Genes
507 and Its Application for Efficient Production of Poly(3-Hydroxybutyrate) and l-Threonine. *Appl.*
508 *Environ. Microbiol.* **87**, e00381-21 (2021).
509 59. Henderson, I. R., Meehan, M. & Owen, P. A novel regulatory mechanism for a novel phase-variable
510 outer membrane protein of Escherichia coli. *Adv. Exp. Med. Biol.* **412**, 349–355 (1997).
511 60. Graveline. Lrp-DNA complex stability determines the level of ON cells in type P fimbriae phase
512 variation. *Mol Microbiol* (2011) doi:10.1111/j.1365-2958.2011.07761.x.
513 61. Wallecha. Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence
514 for a new role of DNA methylation. *J Bacteriol* (2002) doi:10.1128/jb.184.12.3338-3347.2002.
515 62. Stevenson, G. *et al.* Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb
516 gene cluster. *J. Bacteriol.* **176**, 4144–4156 (1994).
517 63. Eguchi, Y. & Utsumi, R. Alkali Metals in Addition to Acidic pH Activate the EvgS Histidine Kinase
518 Sensor in Escherichia coli. *J. Bacteriol.* **196**, 3140–3149 (2014).
519 64. Kern, R., Malki, A., Abdallah, J., Tagourti, J. & Richarme, G. Escherichia coli HdeB Is an Acid
520 Stress Chaperone. *J. Bacteriol.* **189**, 603–610 (2007).
521 65. Yasui, R., Washizaki, A., Furihata, Y., Yonesaki, T. & Otsuka, Y. AbpA and AbpB provide anti-
522 phage activity in *Escherichia coli*. *Genes Genet. Syst.* **89**, 51–60 (2014).

523 66. Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A. & Cheng, X. Structure and Substrate Recognition
524 of the Escherichia coli DNA Adenine Methyltransferase. *J. Mol. Biol.* **358**, 559–570 (2006).
525 67. Maier, J. A. H. & Jeltsch, A. Design and Application of 6mA-Specific Zinc-Finger Proteins for
526 the Readout of DNA Methylation. in *Zinc Finger Proteins: Methods and Protocols* (ed. Liu, J.) 29–41
527 (Springer, 2018). doi:10.1007/978-1-4939-8799-3_3.
528 68. Sera, T. & Uranga, C. Rational design of artificial zinc-finger proteins using a nondegenerate
529 recognition code table. *Biochemistry* **41**, 7074–7081 (2002).
530 69. Nabel, C. S., Manning, S. A. & Kohli, R. M. The Curious Chemical Biology of Cytosine:
531 Deamination, Methylation,and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol.* **7**,
532 20–30 (2012).
533 70. Shapiro, R. & Klein, R. S. The Deamination of Cytidine and Cytosine by Acidic Buffer Solutions.
534 Mutagenic Implications*. *Biochemistry* **5**, 2358–2362 (1966).
535 71. Lindahl, T. & Nyberg, B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid.
536 *Biochemistry* **13**, 3405–3410 (1974).
537 72. O'Brown, Z. K. & Greer, E. L. N6-Methyladenine: A Conserved and Dynamic DNA Mark. in *DNA*
538 *Methyltransferases - Role and Function* (eds. Jeltsch, A. & Jurkowska, R. Z.) 213–246 (Springer
539 International Publishing, 2016). doi:10.1007/978-3-319-43624-1_10.
540 73. Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F. & Ochoa, S. Synthetic polynucleotides and the
541 amino acid code, V*. *Proc. Natl. Acad. Sci.* **48**, 613–616 (1962).
542 74. Sater, M. R. A. *et al.* DNA Methylation Assessed by SMRT Sequencing Is Linked to Mutations in
543 Neisseria meningitidis Isolates. *PLOS ONE* **10**, e0144612 (2015).
544 75. Brunet. Fur-dam regulatory interplay at an internal promoter of the enteroaggregative Escherichia
545 coli Type VI secretion sci1 gene cluster. *J Bacteriol* (2020) doi:10.1128/JB.00075-20.
546 76. Mehling. A Dam methylation mutant of Klebsiella pneumoniae is partially attenuated. *FEMS*
547 *Microbiol Lett.* (2007) doi:10.1111/j.1574-6968.2006.00581.x.
548 77. McIntyre, A. B. R. *et al.* Single-molecule sequencing detection of N6-methyladenine in microbial
549 reference materials. *Nat. Commun.* **10**, 579 (2019).

550 78. Oxford Nanopore Technologies. Tombo Documentation Modified Base Detection.
551 https://nanoporetech.github.io/tombo/modified_base_detection.html.
552 79. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
553 polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012).
554 80. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
555 81. Santos-Zavaleta, A. *et al.* RegulonDB v 10.5: tackling challenges to unify classic and high
556 throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.* **47**, D212–D220
557 (2019).
558 82. Wang, H.-Q., Tuominen, L. K. & Tsai, C.-J. SLIM: a sliding linear model for estimating the
559 proportion of true null hypotheses in datasets with dependence structures. *Bioinforma. Oxf. Engl.*
560 **27**, 225–231 (2011).
561 83. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by
562 microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
563 84. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful
564 Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
565 85. Bolstad, B. M., Irizarry, R. A., Åstrand, M. & Speed, T. P. A comparison of normalization methods
566 for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193
567 (2003).
568 86. Oksanen, J. *et al. vegan: Community Ecology Package*. (2022).
569 87. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The*
570 *Innovation* **2**, 100141 (2021).
571 88. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological
572 Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
573 89. Carlson, M. org.EcK12.eg.db: Genome wide annotation for E coli strain K12. (2019).
574



575
576 **Figure 1. Genome-wide 6mA methylation in *E. coli* K-12 substr. MG1655.** The origin of replication
577 (green), IS elements (purple) and cryptic prophages (orange) are shown for reference. (A) Percent
578 methylation at GATC sites. Each point represents one GATC site and points are colored to reflect
579 percent methylation from 0% (blue, inside) to 100% (red, outside). (B-C) Differentially methylated GATC
580 sites in evolved WT (A) and MMR- (B) clones. Red bars are sites with increased methylation and blue
581 bars are decreased methylation, and the height of the bar reflects the difference in percent methylation
582 between the ancestor and the evolved clones. Each ring ranges from -25% to 25%.

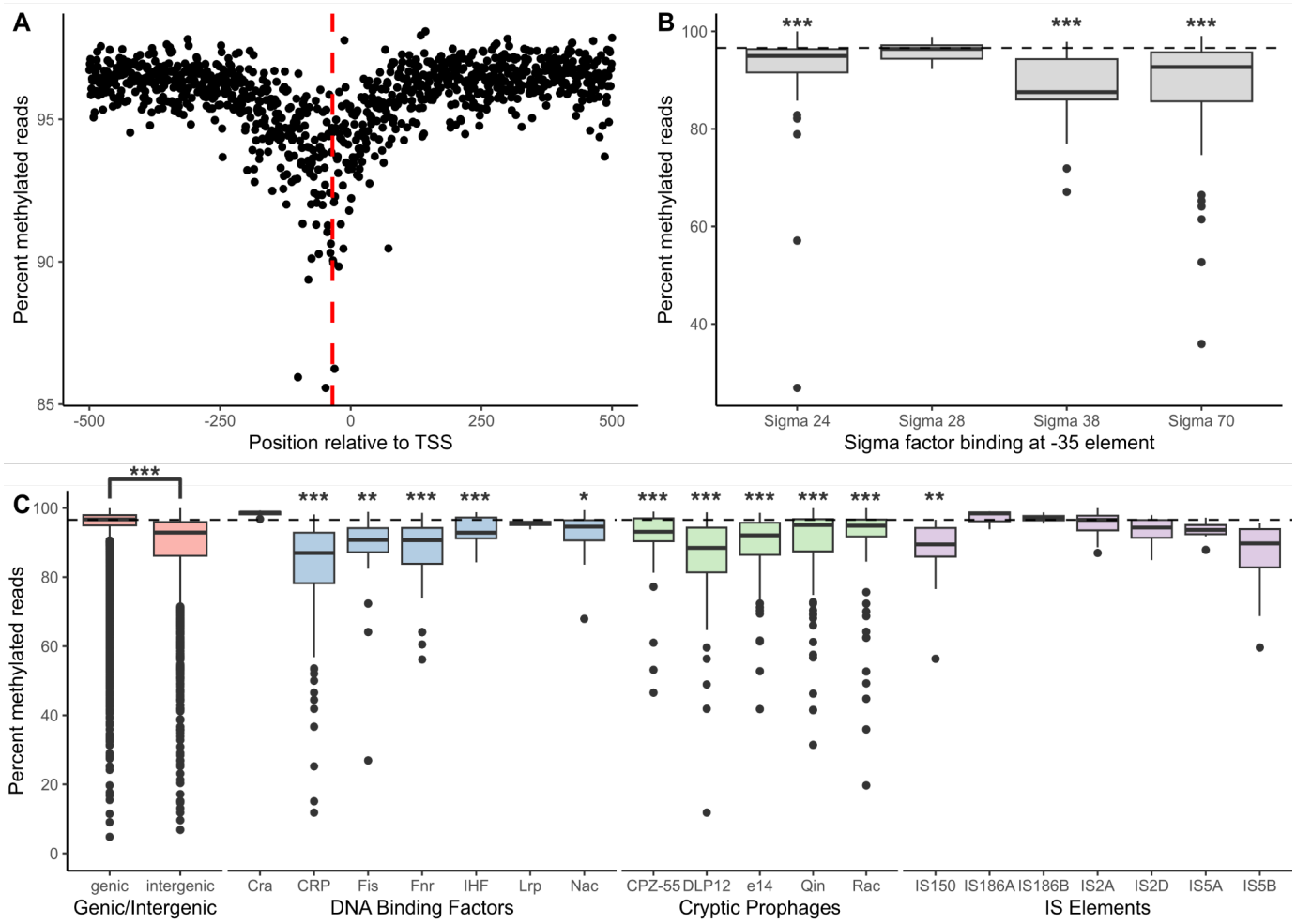


Figure 2. Genome-averaged methylation decreases around core promoter elements. (A) Median methylation at GATC sites aligned relative to the transcription start sites (TSS) for all promoters. Points represent the median methylation for all GATC sites at that position. The -35 position relative to the 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. (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.

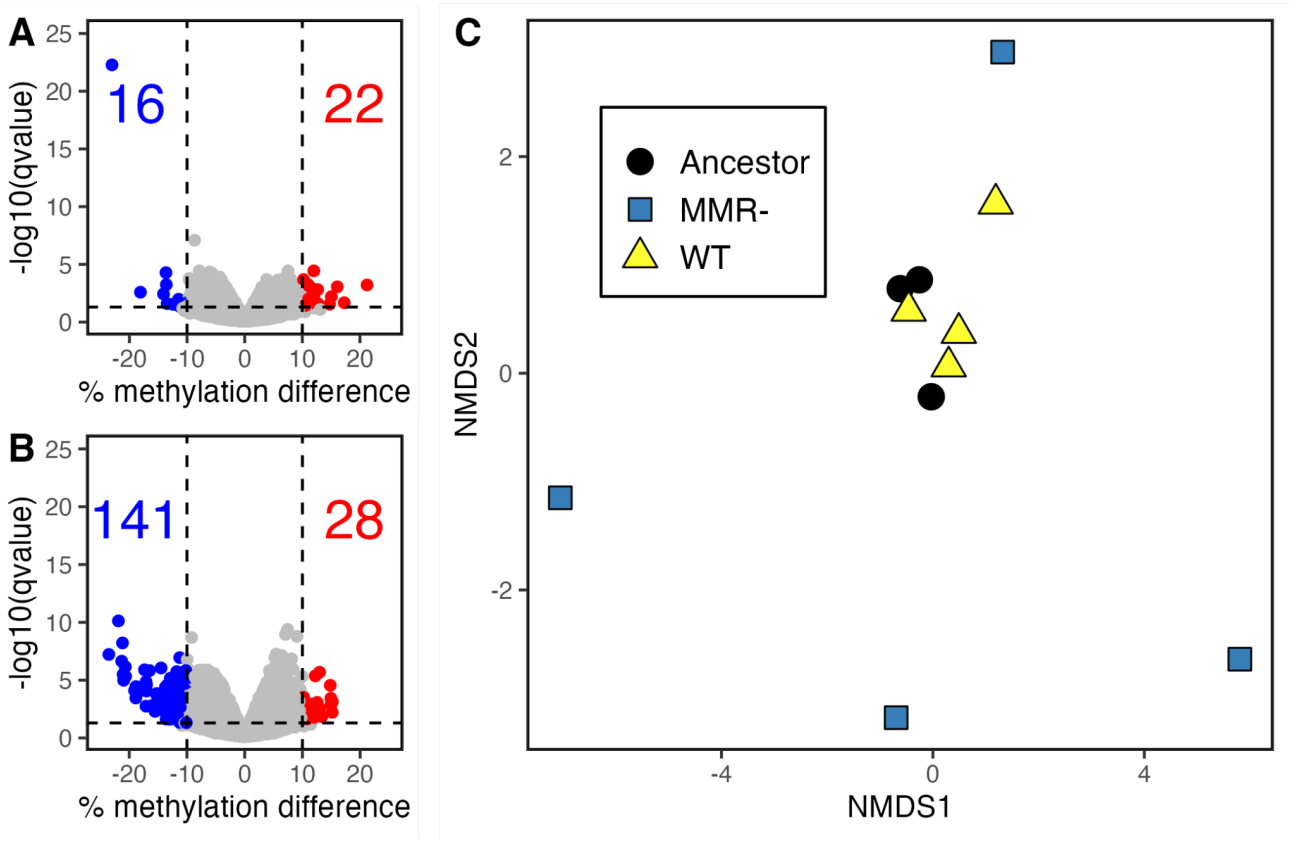


Figure 3. Differentially methylated GATC sites between the ancestor and experimentally evolved (A) WT clones or (B) MMR- clones. Differential methylation was called for WT and MMR- clones compared to the ancestor strain using methylKit. Differential methylation was determined for each site as a percent 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 β-values and ordination was performed with NMDS. Each point represents one clone.

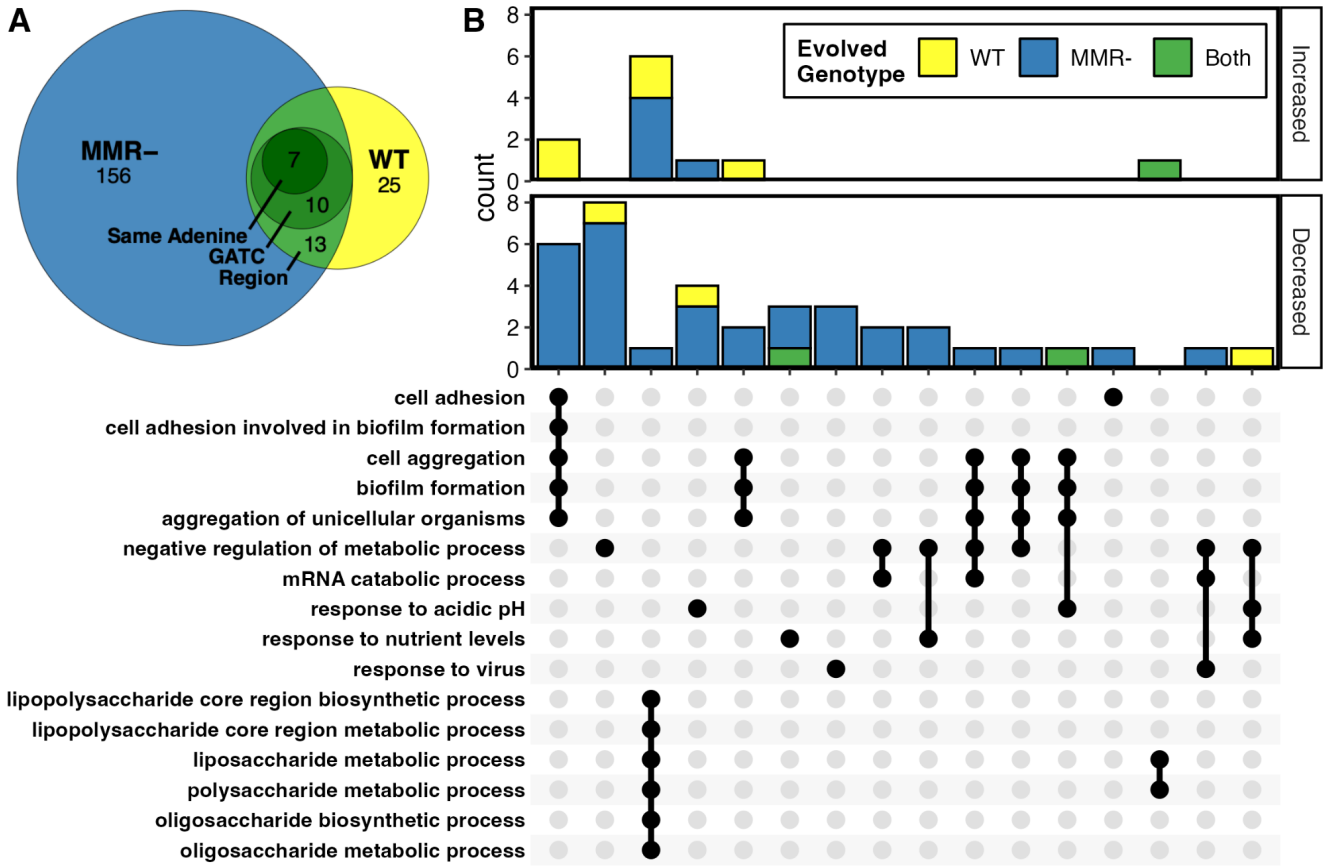


Figure 4. Gene ontology enrichment analysis of differentially methylated GATC sites. (A) Intersect between differentially methylated sites in evolved WT and MMR- genotypes. Sites in the same region are located in the same gene or the same intergenic region. (B) UpSet plot of significantly enriched gene 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.

Table 1. Overrepresentation of differentially methylated adenines in protein-DNA interaction motifs.

		Evolved WT				Evolved MMR-			
Feature Type	**# GATC sites in ancestor**	**Increased methylation**	**p-value**	**Decreased methylation**	**p-value**	**Increased methylation**	**p-value**	**Decreased methylation**	**p-value**
Transcriptional regulator binding sites	313	7	0.095	**12***	**< 0.001**	**11^**	**0.014**	**58***	**< 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***	**< 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***	**0.043**
Sigma 70 binding sites	86	1	0.694	2	0.313	3	0.254	**12***	**< 0.001**
Genes	35595	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***	**0.003**	0	1.000	**6***	**0.011**
IS elements	283	**10^**	**0.002**	**10***	**0.001**	**11^**	**0.007**	**18***	**0.014**
Cryptic prophages	795	**15^**	**0.038**	11	0.249	**24^**	**0.001**	**62***	**< 0.001**
No Feature	1191	**35^**	**< 0.001**	**30***	**< 0.001**	**44^**	**< 0.001**	**134***	**< 0.001**

P-values were calculated by one-tailed Fisher's exact test followed by BH correction. ^ = significant number of sites with increased methylation. * = significant number of sites with decreased methylation.

613

Table 2. Number of mutations in GATC sites in evolved clones.

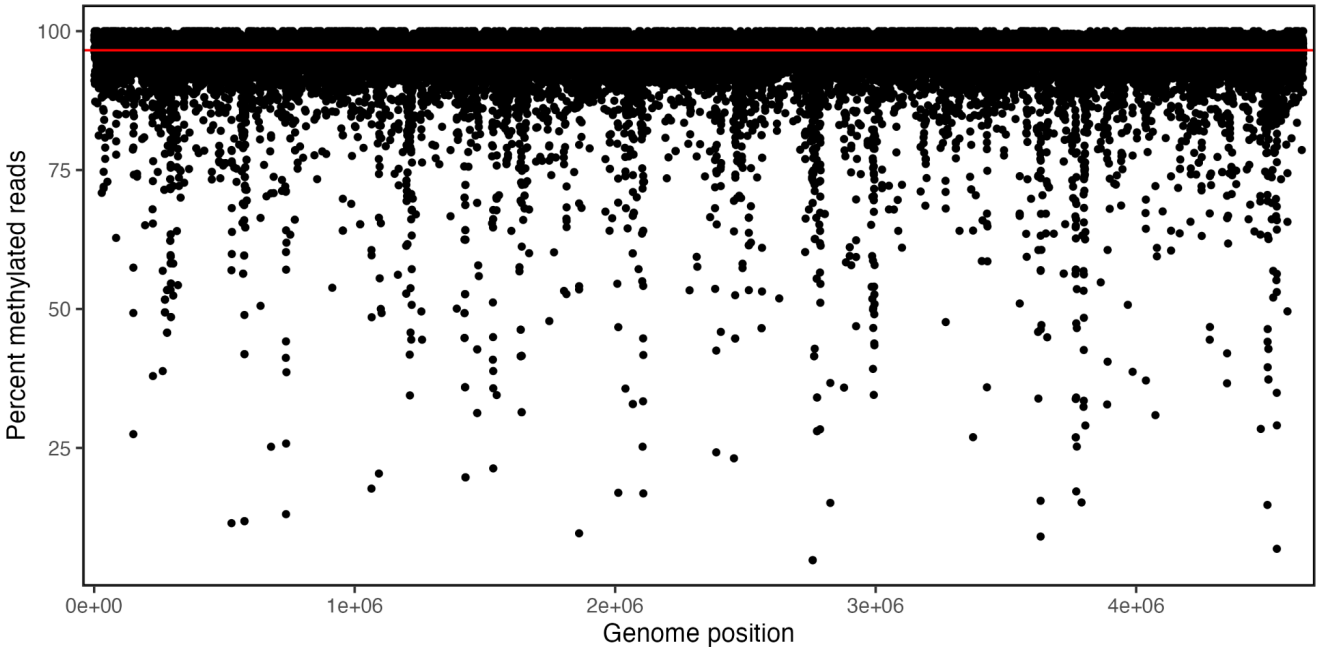
Background	Population*	Clones^	Mutations	GATC Gain/Loss	Percent GATC Mutated (This Study)	Percent GATC Mutated (Lee et al, 2012)
MMR	113		257	13	0.051	
		113-1	121	3	0.025	
		113-2	96	3	0.031	
	126		46	1	0.022	
		126-2	313	11	0.035	
		126-4	126	3	0.024	
	Total		959	34	0.035 (0.028 - 0.040)	0.035 (0.028 - 0.043)
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)	0.043 (0.025 - 0.067)

*Mutations shared between clones and occurred before subpopulation divergence

^Mutations unique to clones and occurred after subpopulation divergence

614

615



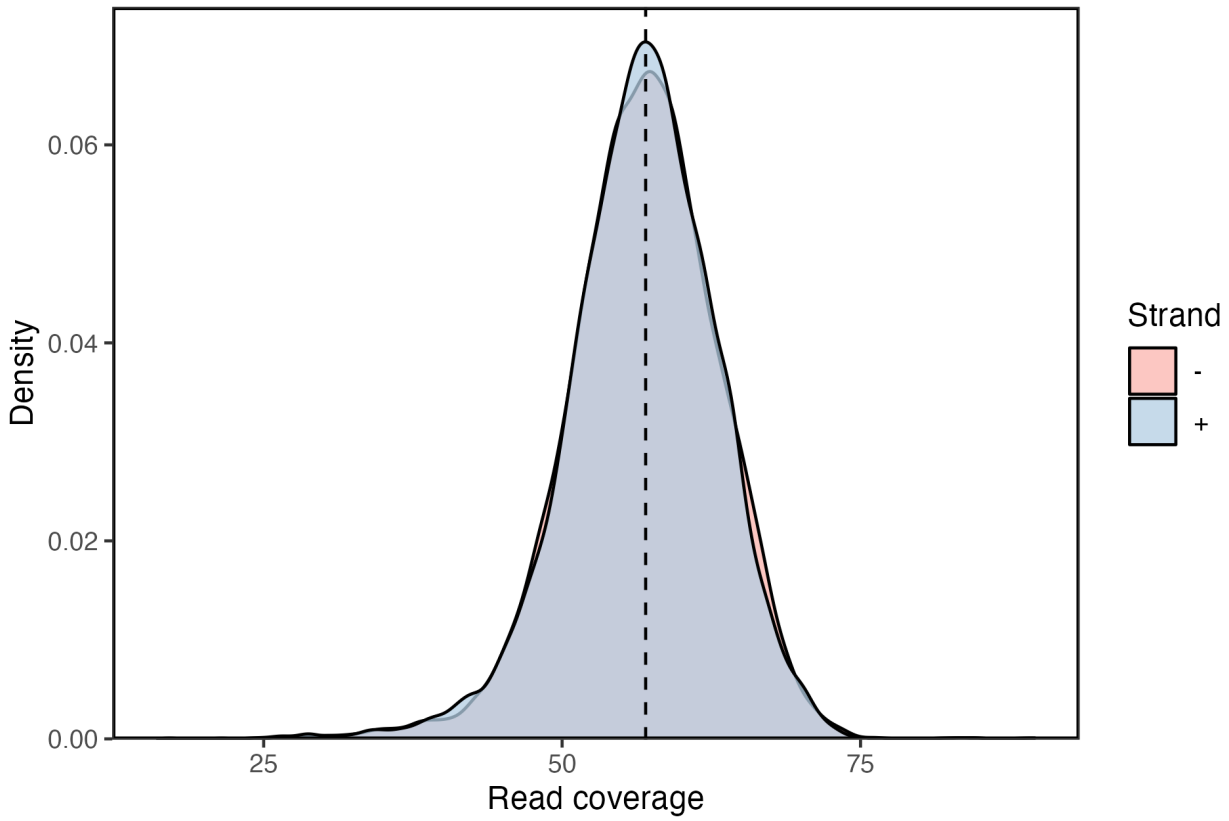
616

617 **Figure S1. 6mA methylation across the MG1655 genome.** Each point represents the mean percent

618 methylation of three sequencing replicates of the K-12 MG1655 ancestor strain at one GATC site. The

619 genome-wide median methylation of 96.6% is shown for reference (red line).

620



621

622 **Figure S2. Sequencing coverage of GATC sites by Nanopore sequencing.** Coverage distributions

623 for each strand are shown separately, and both distributions have the same median of 57 reads

624 (dashed line).

625

626

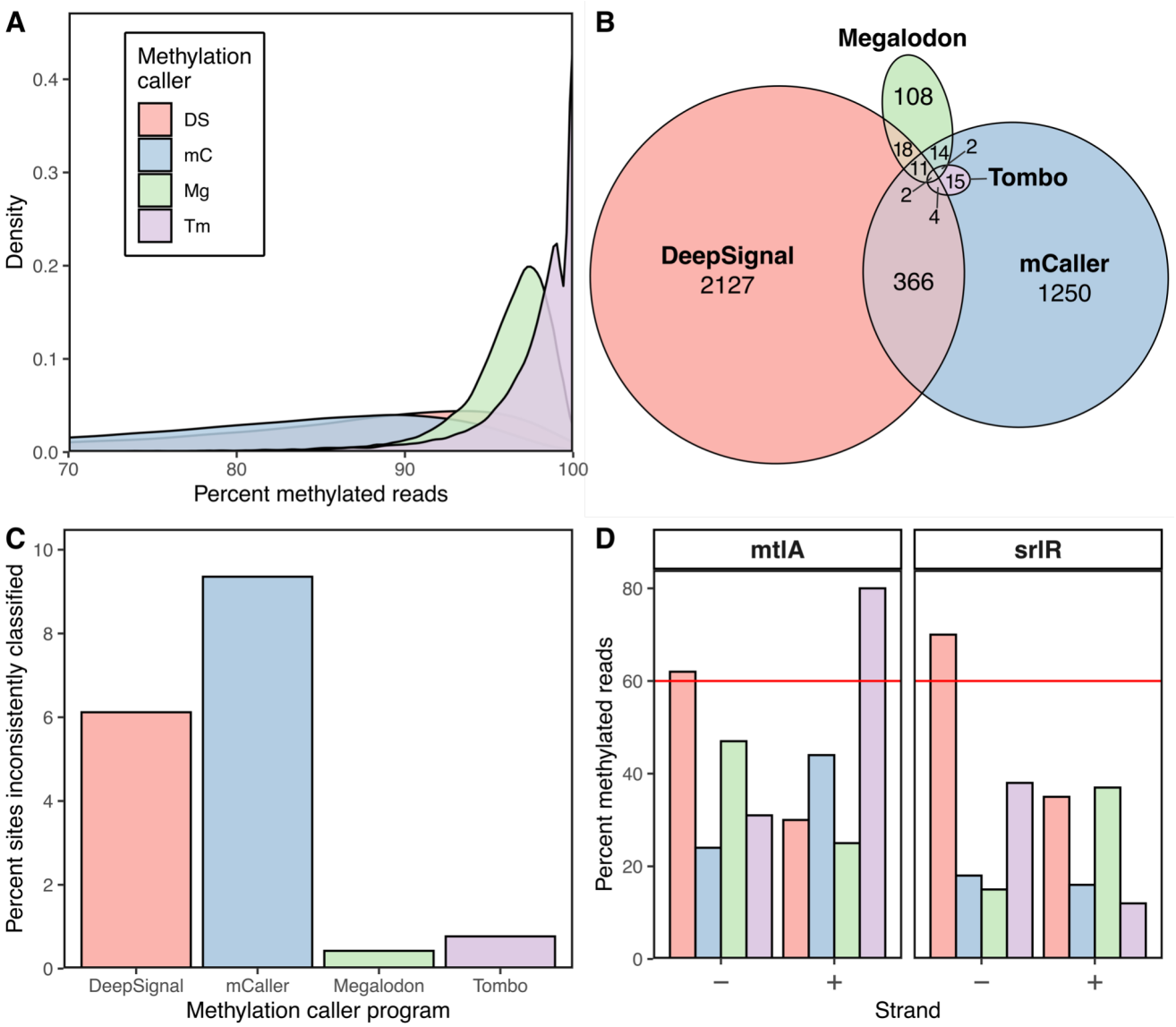


Figure S3. Comparison of bioinformatics programs on identifying methylated reads at GATC sites. (A) Distribution of percent of reads classified as methylated at all GATC sites in the genome for each caller. Dashed lines represent median percent methylation across all GATC sites for, from left to right, mCaller (83%), DeepSignal (87%), Megalodon (97%), and Tombo (99%). (B) Overlap between GATC sites classified as undermethylated by each program. Each area is labeled with the number of GATC sites. (C) Percentage of methylated GATC sites inconsistently classified as methylated across WT replicates for each program. (D) Percentage of reads classified as methylated by each program at GATC sites previously reported as <60% methylated.

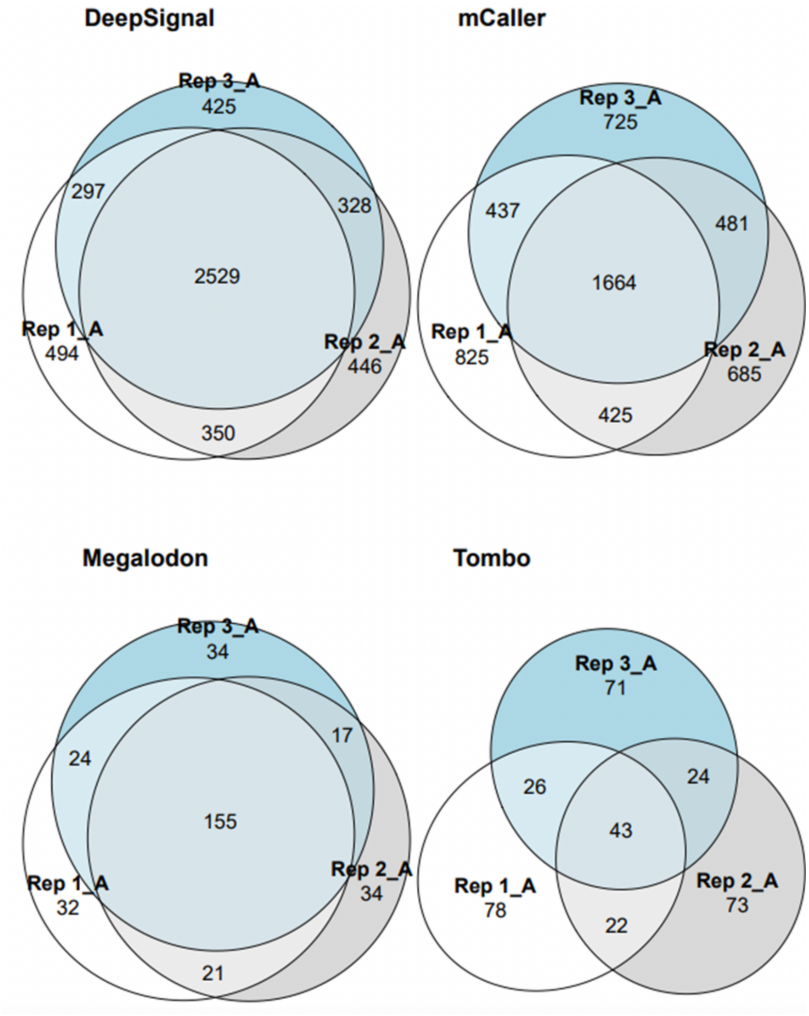
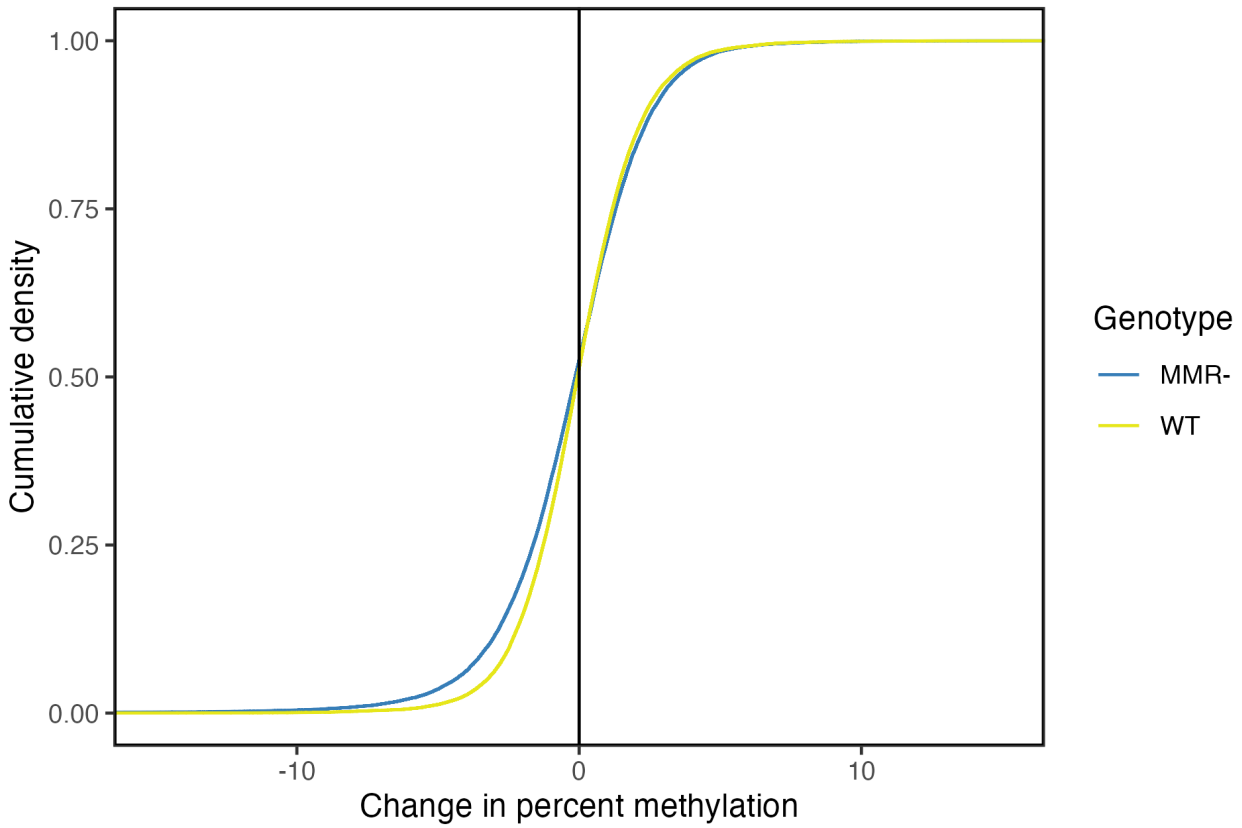


Figure S4. Overlap in sites identified as hypomethylated between three sequencing replicates for each caller. Each area is labeled with the number of GATC sites.



641

642 **Figure S5. Distribution of changes in per-site methylation between evolved genotypes and their**

643 **isogenic ancestor.** For each GATC site, the difference between average percent methylation in the

644 evolved genotype and in the ancestor was calculated, and values were used to calculate empirical

645 cumulative density functions.

646

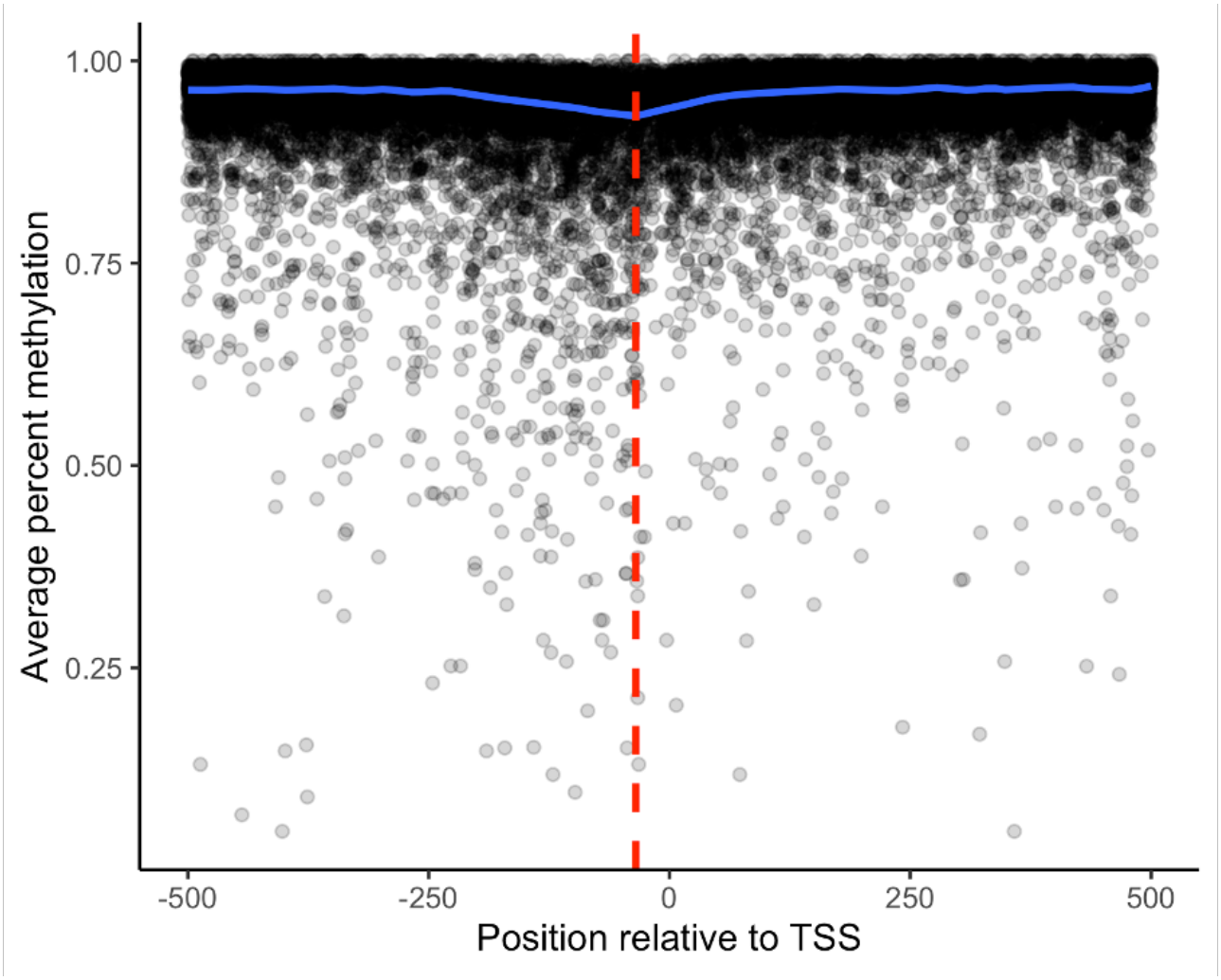


Figure S6. Methylation of all GATC sites relative to the transcription start site (TSS). Each point represents a single GATC site. The smoothed median was calculated by additive quantile regression smoothing (blue line), and the -35 position is shown for reference (dashed red line).

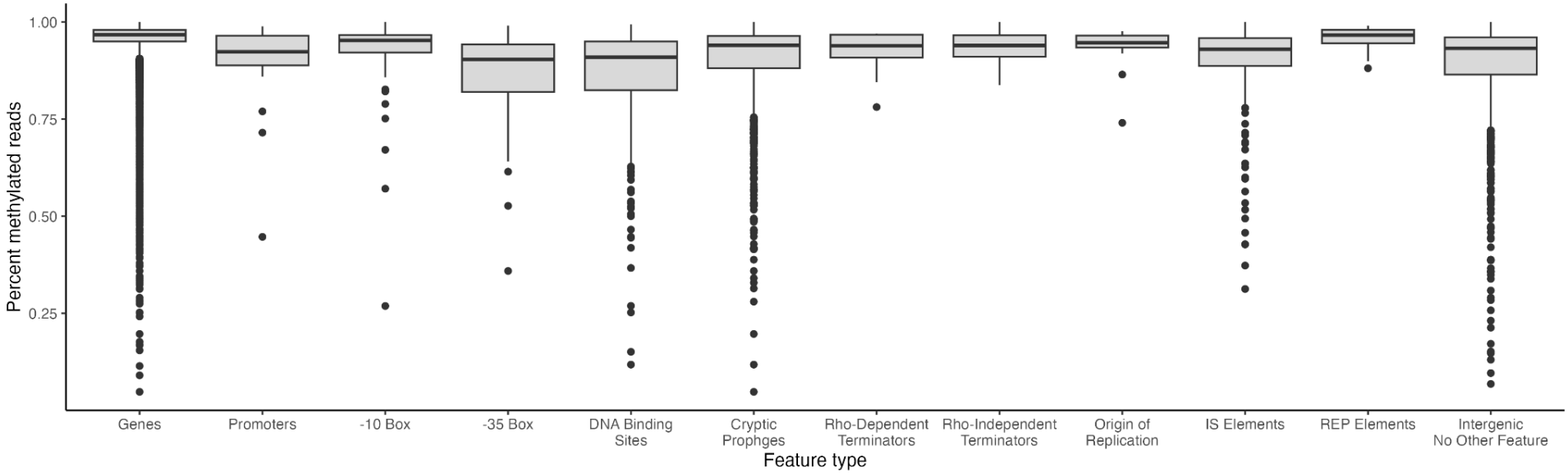


Figure S7. Average methylation at various genomic features. Boxplots show the median, first quartile, and third quartile of the percent methylated reads.