bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methylation by multiple type I restriction modification systems avoids influencing gene regulation in uropathogenic *Escherichia coli*

Kurosh S. Mehershahi[1] and Swaine L. Chen[1,2,*]

[1]NUHS Infectious Diseases Translational Research Programme, Department of Medicine, Division of Infectious Diseases, Yong Loo Lin School of Medicine, Singapore

[2]Laboratory of Bacterial Genomics, Genome Institute of Singapore, Singapore

[*]Corresponding author.
Email: slchen@gis.a-star.edu.sg (SLC)

Running title: No regulatory impact of type I methylation in UPEC

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

DNA methylation is a common epigenetic mark that influences transcriptional regulation, and therefore cellular phenotype, across all domains of life, extending also to bacterial virulence. Both orphan methyltransferases and those from restriction modification systems (RMSs) have been co-opted to regulate virulence epigenetically in many bacteria. However, the potential regulatory role of DNA methylation mediated by archetypal Type I systems in *Escherichia coli* has never been studied. We demonstrated that removal of DNA methylated mediated by three different *Escherichia coli* Type I RMSs in three distinct *E. coli* strains had no detectable effect on gene expression or growth in a screen of 1190 conditions. Additionally, deletion of the Type I RMS EcoUTI in UTI89, a prototypical cystitis strain of *E. coli*, which led to loss of methylation at >750 sites across the genome, had no detectable effect on virulence in a murine model of ascending urinary tract infection (UTI). Finally, introduction of two heterologous Type I RMSs into UTI89 also resulted in no detectable change in gene expression or growth phenotypes. These results stand in sharp contrast with many reports of RMSs regulating gene expression in other bacteria, leading us to propose the concept of “regulation avoidance” for these *E. coli* Type I RMSs. We hypothesize that regulation avoidance is a consequence of evolutionary adaptation of both the RMSs and the *E. coli* genome. Our results provide a clear and (currently) rare example of regulation avoidance for Type I RMSs in multiple strains of *E. coli*, further study of which may provide deeper insights into the evolution of gene regulation and horizontal gene transfer.

Author summary

DNA methylation is perhaps the most common epigenetic modification, and it is commonly associated with gene regulation (in nearly all organisms) and virulence (particularly well studied in bacteria). Regarding bacterial virulence, the current DNA methylation literature has focused primarily on orphan methyltransferases or phasevariable restriction modification systems (RMSs). Interestingly, no reports have studied the potential regulatory role of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

first RMS discovered, the Type I RMS EcoKI. We used transcriptomics, Phenotype Microarrays, and a murine model of urinary tract infection to screen for functional consequences due to Type I methylation in three unrelated strains of *E. coli*. Remarkably, we found zero evidence for any epigenetic regulation mediated by these Type I RMSs. Thus, these Type I RMSs appear to function exclusively in host defense against incoming DNA (the canonical function of RMSs), while the methylation status of many hundreds of the corresponding recognition sites has no detectable impact on gene expression or any phenotypes. This led us to the concept of "regulation avoidance" by such DNA methyltransferases, which contrasts with the current literature on bacterial epigenetics. Our study hints at the existence of an entire class of regulation avoidant systems, which provides new perspectives on methylation-mediated gene regulation and bacterial genome evolution.

Keywords: Restriction-modification systems; epigenetics; DNA methylation; uropathogenic *Escherichia coli*; Single molecule real-time sequencing

Introduction

Regulation and response are fundamental aspects of cellular life. Epigenetics in the context of regulation refers to heritable changes in gene expression without alterations to the primary DNA sequence of an organism (1). This rapidly growing field encompasses an expanding list of biochemical incarnations spanning the entire central dogma (DNA and RNA modifications; chromatin remodelling; noncoding RNAs; and prion proteins) that adds additional layers of complexity and regulatory potential on top of primary DNA sequences (2–5). Epigenetics thus allows a clonal population to generate phenotypic heterogeneity, which can serve specific biological functions (6,7). Particularly for pathogens (and especially facultative pathogens), epigenetics is one way to implement rapid responses to new environments such

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

as different host niches and immune pressures, which in turn may enhance survival and virulence (8,9).

DNA methylation represents the most extensively studied epigenetic mechanism in all domains of life (2,10–12). In eukaryotes, DNA methylation has been demonstrated to play an important role in differentiation, development and disease (including cancer) (13,14). Bacterial DNA methylation, in particular, has a special place in the collection of epigenetic implementations, as its role “outside genetics” in mediating phage resistance was discovered and characterized prior to the modern usage of the term epigenetics (15).

Bacterial DNA methylation occurs most commonly on adenine (N6-methyladenine, 6mA), but also on cytosine (C5-methylcytosine, 5mC and N4-methylcytosine, 4mC) nucleotides (16). The majority of known bacterial DNA methyltransferases belong to two general categories, restriction modification system (RMS)-associated and orphan methyltransferases (4). RMSs have been characterized as bacterial innate immune systems, using methylation of specific motifs to differentiate between self and non-self DNA (15). RMSs can be found in 90% of all sequenced prokaryotic genomes, with 80% possessing multiple systems, hinting at a diverse reservoir of potential epigenetic information (11). RMSs are classified based on co-factor requirement, subunit composition, cleavage pattern, and mechanism into 4 types (I to IV) (17–20). Type I RMSs were the first discovered and are multi-subunit enzyme complexes consisting of three proteins: a methyltransferase (typically denoted HsdM), an endonuclease (HsdR), and a sequence recognition protein (HsdS). Type I RMSs recognise bipartite sequences (for example, 5’-AAC(N_6)GTGC-3’, where N=A,T,G or C) with each HsdS subunit possessing two highly variable target recognition domains (TRDs) which specify the two halves of the bipartite motif (19).

Beyond immunity, RMS-associated methyltransferases are also known to regulate gene expression. A dramatic example of such regulation is found in phasevariable Type I and Type III RMSs. Phase variation of these RMSs results in rapid and reversible changes to

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

methylation patterns, resulting in coordinated changes to the expression of distinct sets of genes (regulons), hence the term “phasevariable regulons” or phasevarions (8,21). Phasevarions can be identified by known markers of phase variation, such as simple sequence repeats (SSRs) within promoters or gene bodies and *hsdS* genes flanked by inverted repeats. Systematic surveys have identified 13.8% Type I and 17.4% Type III RMSs as being potentially phase variable (22–24). Several clinically relevant pathogens such as *Helicobacter pylori* (25,26), *Neisseria meningitidis* (27), *Neisseria gonorrhoeae* (28,29), *Haemophilus influenzae* (30,31), *Moraxella catarrhalis (32,33)*, and *Kingella kingae* (34) harbor such phase variable RMSs. An extreme example is found in *Streptococcus pneumoniae*, where some strains carry a phase variable Type I RMS which can result in up to six distinct methylation specificities, with different methylation patterns specifically associated with either invasive disease or nasopharyngeal carriage *in vivo (35–37).*

In contrast, orphan methyltransferases, as their name suggests, are methylating enzymes without a functional cognate endonuclease. Unlike RMSs, orphan methyltransferases are phylogenetically more conserved across genera/families (11,38,39) and even essential in certain bacteria (40–42). Orphan methyltransferases are not involved in host defense; instead, they play key housekeeping regulatory roles, ranging from DNA replication, mismatch repair, controlling transposition, and gene expression (43). Regarding gene expression, mutation of these orphan methyltransferases can have large effects on the transcriptome, affecting hundreds to thousands of genes (44–49); not surprisingly, therefore, mutation of orphan methyltransferase can also lead to defects in virulence (50–55). In some cases, there exist locus-specific roles of methylation in virulence factor regulation, such as for P pili and the autotransporter antigen 43 (Agn43) (56,57). The two most well-studied orphan methyltransferases are Dam (DNA adenine methyltransferase) in gamma-proteobacteria and CcrM (cell cycle regulated DNA methyltransferase) in alpha-proteobacteria (58).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Therefore, DNA methylation can generally affect both gene expression and phenotypes, of which virulence is an especially interesting case in pathogens. Interestingly, there has been a relative dearth of such studies on the archetypal Type I RMSs, which are typically non-phase variable and retain both restriction and modification functions. We used single molecule sequencing to characterize the Type I RMS in a uropathogenic *Escherichia coli* strain, UTI89. We verified that this Type I RMS was functional for restriction of incoming transformed DNA. Surprisingly, we could find no effect on gene expression or virulence when this RMS was removed, despite >700 DNA bases changing in methylation state. Similar results were found when we removed the Type I RMS from two other *E. coli* strains, MG1655 and CFT073. Intrigued by these finding, we replaced the native UTI89 Type I RMS with the Type I RMS from MG1655 and CFT073; in both cases, re-installing methylation at ~400-600 distinct sites in the genome also had no effect on gene expression or on a broad panel of bacterial growth phenotypes. These data suggest that, in *E. coli,* these canonical Type I systems are purely host defense mechanisms devoid of any secondary regulatory functions in host physiology and virulence. This is the first reported example of "regulation avoidance" for any DNA methylation system in bacteria. The fact that three distinct methylation specificities all demonstrate this regulation avoidance raises further speculations about the evolution and plasticity of both the genome and the Type I RMSs in *E. coli*.

Results

Uropathogenic *Escherichia coli* UTI89 possesses a functional Type I Restriction Modification System

The clinical cystitis strain UTI89 encodes a Type I RMS at the *hsdSMR* locus (59,60). PacBio single molecule real time (SMRT) sequencing identified 3 methylation motifs in UTI89: 5'-GATC-3', 5'-CCWGG-3', and 5'-CCA(N_7)CTTC-3' (S1 Table). The first two motifs are well known in *E. coli* as the methylation sites of orphan methyltransferases Dam and Dcm, respectively (12,61). The third motif, a bipartite N6-methyladenine (m6A) motif, is

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

typical for a Type I RMS (19). Deletion of the *hsdSMR* locus in UTI89 resulted in loss of adenine methylation only at the 5'-CCA(N_7)CTTC-3' motif, confirming the specificity of this RMS (S1 Table).

To verify that the UTI89 *hsdSMR* locus was functional for restriction of incoming DNA, we performed a classic plasmid transformation efficiency assay (62). A plasmid containing the 5'-CCA(N_7)CTTC-3' motif was isolated from the UTI89*ΔhsdSMR* strain (which does not methylate this motif); this plasmid was transformed less efficiently into wild type (wt) UTI89 than into an otherwise isogenic UTI89*ΔhsdSMR* strain (Fig 1A). Moreover, as seen with other Type I RMSs, addition of a second motif further reduced the efficiency of transformation into wt UTI89 (Type I RMSs cleave DNA when translocation is stalled due to collision between adjacent Type I complexes (63)). When plasmids were isolated from wt UTI89 (which methylates the 5'-CCA(N_7)CTTC-3' motif), no difference in efficiency was seen between transformations into UTI89 and UTI89*ΔhsdSMR* (Fig 1A). Thus, uropathogenic *E. coli* UTI89 possesses a functional Type I RMS, with a formal name of RM.EcoUTI and a specificity of 5'-CCA(N_7)CTTC-3' (64). We hereafter refer to this system as EcoUTI.

EcoUTI Type I methylation has no detectable role during *in vivo* urinary tract infection

DNA methylation is widely recognized to influence gene expression and virulence in multiple bacterial pathogens (51,58,65). EcoUTI methylates 754 sites in the UTI89 genome; we therefore hypothesized that complete removal of this system would affect both gene expression and virulence. The UTI89 and UTI89*ΔhsdSMR* strains had no difference in growth in both rich and minimal media nor in Type I pilus expression *in vitro* (data not shown). We then tested virulence using a transurethral murine model for ascending urinary tract infection (UTI) (66). Competitive co-infections in 6-8 week old C3H/HeN mice using equal mixtures of wild type UTI89 and UTI89*ΔhsdSMR* showed no competitive advantage for either strain at 1 or 7 days post-infection (dpi) in either bladders or kidneys (Fig 1B).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Comparison of bacterial loads from single infections also showed no significant difference between the two strains (S1 Fig).

EcoUTI Type I methylation has no effect on gene expression or growth in multiple conditions

The lack of an *in vivo* infection phenotype, despite the complete removal of EcoUTI methylation, contrasts with reports on multiple other methylation systems (including RMSs) that impact bacterial phenotypes (47,67–71). We therefore asked whether changing methylation could alter any gene expression using RNA-seq. When comparing UTI89 with the otherwise isogenic UTI89*ΔhsdSMR* strain, in both log and stationary phase in rich media, only 7 genes were significantly changed in expression in the mutant: the (deleted) *hsdSMR* and UTI89_C5052 genes were downregulated, the two genes flanking the *hsdSMR* locus were upregulated, and a single transposase gene was upregulated. In the UTI89 genome, UTI89_C5052 is a hypothetical gene annotated in the intergenic region between *hsdM* and *hsdR*, so it was deleted with the entire *hsdSMR* locus. Altered expression of the two flanking genes was connected to the use of an antibiotic resistance cassette to knock out *hsdSMR* (see methods). Targeted qRT-PCR on the transposase gene was unable to validate the change seen in RNA-seq (Fig 1C, S2 Fig, S2 and S3 Tables). Therefore, we could find no change in gene expression that was attributable to loss of Type I methylation.

Although RNA-seq is a powerful tool for assaying gene expression, it only tests expression in one experimental growth condition. To test a broader range of growth conditions, we used Biolog Phenotype Microarrays (PM) (72). PM plates 1-20 represent 1190 different nutrients, toxins, antibiotics, inhibitors, and other conditions. Relative to wt UTI89, the UTI89*ΔhsdSMR* strain had gained resistance to the aminoglycoside antibiotic paromomycin (PM12, wells C01-02), which was consistent with the use of a kanamycin resistance cassette to knock out the *hsdSMR* locus. (Fig 1D, S4 Table). UTI89*ΔhsdSMR* had also gained resistance to the formazan dye Iodonitrotetrazolium (INT) violet (PM19, well D07). There are four wells

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

containing different concentrations of INT on PM19 (D05 - D08); well D07 is the second-lowest concentration, and the other three wells were not called as different from wt UTI89. In particular, at the two highest concentrations (wells D05 and D06) there was very little difference in the growth curves between UTI89 and UTI89*ΔhsdSMR.* We therefore suspect that this was a false positive phenotype (Fig 1D, S4 Table). We thus find, strikingly, no phenotypic difference attributable to the loss of methylation.

Loss of Type I methylation in two other *E. coli* strains also has no effect on gene expression or growth phenotypes

We next asked whether the lack of any detectable gene expression or growth phenotype changes was unique to UTI89 and the EcoUTI methylation system. *E. coli* strain MG1655 is a well-studied, lab-adapted K12 strain, while CFT073 is a commonly used pyelonephritis strain; both also carry a single Type I RMS (EcoKI and EcoCFTI, respectively). EcoKI methylates the 5'-AAC(N_6)GTGC-3' motif, while EcoCFTI is predicted to methylate the 5'-GAG(N_7)GTCA-3' motif (59,73). Plasmid transformation efficiency assays confirmed that both EcoKI and EcoCFTI also were functional as restriction systems for incoming DNA, targeting the previously reported motifs (S3 Fig). In RNA-seq experiments, besides the expected change in expression of the *hsdSMR* genes themselves and the flanking genes affected by insertion of the antibiotic resistance cassette, MG1655*ΔhsdSMR* had 4 genes upregulated during log phase, all encoded within a predicted cryptic prophage and validated by qRT-PCR (Fig 2A, S4A Fig, S2 and S3 Tables). However, independently generated clones of the MG1655*ΔhsdSMR* strain did not have a similar upregulation of these four prophage genes when tested by qRT-PCR; we therefore consider the expression changes in these four genes an artefact of that single strain, possibly due to a second site mutation during generation of the knockout (data not shown). The CFT073*ΔhsdSMR* strain had limited changes in gene expression relative to the wild type CFT073, including the *hsdSMR* genes, and flanking gene *yjiW* (polar effect). Targeted qRT-PCR on *malK* and *lamB* genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was unable to validate the changes seen in RNA-seq, confirming these as false positives (Fig 2B, S4B Fig, S2 and S3 Tables).

Using the Biolog PM, we again found no phenotypic differences between either (i) MG1655 and MG1655*ΔhsdSMR* or (ii) CFT073 and CFT073*ΔhsdSMR*; except for gained antibiotic resistance due to kanamycin resistance cassette used to knock out the *hsdSMR* locus (S4 Table, S4C and S4D Figs). Thus, in three distinct *E. coli* strains, each encoding Type I RMSs with different specificities, deletion of the entire RMS had no detectable effect on gene expression or any growth phenotypes we measured.

Switching Type I methylation systems in UTI89 also does not affect gene expression or any growth phenotypes

Type I RMSs can be highly polymorphic among different strains of the same species. In many cases, this is thought to be due to both mutation and recombination, possibly driven by diversifying selection (74,75). Furthermore, the whole *hsdSMR* locus need not be recombined; Type I RMSs are classified into 5 families (designated A through E) based on the similarity of the *hsdM* and *hsdR* genes, enabling intra-family genetic complementation with divergent *hsdS* genes (19). As *hsdS* encodes the specificity determinant, mutation or recombination of just the *hsdS* gene is sufficient to alter the methylation and restriction specificity of the entire system (35,76). The RMSs in UTI89 and MG1655 are both subclassified as Type IA and demonstrate this latter relationship; the *hsdM* and *hsdR* genes in these two strains have very similar sequences (99% identical), while the *hsdS* genes are only 45.4% identical and direct distinct specificities (Fig 3A). The Type I RMS in CFT073, on the other hand, is a Type IB RMS with <40% identity to the Type IA system in all three genes (Fig 3A).

To analyze the effect of changing Type I methylation specificity, we created two derivatives of UTI89: one where the *hsdS* gene was replaced by the MG1655 *hsdS* allele (UTI89 $hsdS^{MG1655}$), and one where the entire *hsdSMR* locus was replaced by the EcoCFTI locus

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(UTI89 *hsdSMR*^{CFT073}). As a control, we also re-inserted the UTI89 *hsdS* allele (UTI89 *hsdS*^{UTI89}) using the same cloning strategy as that used to make UTI89 *hsdS*^{MG1655} (i.e. UTI89 *hsdS*^{MG1655} and UTI89 *hsdS*^{UTI89} share the same parental strains and have undergone the same cloning steps). PacBio SMRT sequencing confirmed that plasmids isolated from each of these strains methylated only the expected motif (data not shown). Plasmid transformation efficiency assays also showed that all 3 "methylation-switch" strains generated indeed had functional Type I RMSs with specificities as expected based on the encoded *hsdS* allele (i.e. UTI89 *hsdS*^{MG1655}, UTI89 *hsdSMR*^{CFT073}, and UTI89 *hsdS*^{UTI89} restricted only plasmids carrying the EcoKI, EcoCFTI, and EcoUTI motif respectively, in a methylation-dependent manner) (Fig 3B).

These strains introduce two different Type I methylation specificities in the UTI89 genome, accounting for 641 (MG1655) and 423 (CFT073) predicted newly methylated sites (Fig 4A, S5 Table). All together, the three RMSs we examined account for 1818 distinct methylation sites that would vary in methylation status among the strains we studied. The number of intergenic methylation sites in UTI89, which are expected to be more likely to affect gene regulation (46,77–80), are comparable for the three Type I RMSs (31-40 intergenic sites) (S5 Table). Using RNA-seq, we again found no gene expression changes for any of the "methylation-switch" strains compared with the parental wt UTI89, except for one hypothetical gene that was close to both the FDR and fold-change cutoffs (Fig 4C and 4D, S5 Fig, S2 and S3 Tables). Furthermore, Biolog PM also showed no significant changes attributable to Type I methylation in UTI89 *hsdS*^{MG1655} (S4 Table, S6 Fig). Notably, there was no difference in INT resistance between UTI89 *hsdS*^{MG1655} and wt UTI89 at any concentration, consistent with our previous interpretation that the single difference identified between UTI89 and UTI89*ΔhsdSMR* was a false positive (Fig 1D, S4 Table).

As a final functional test, we used several *in vitro* and *in vivo* assays for virulence. Changing the methylation system had no effect on growth rate in rich or minimal media or on virulence-related assays such as motility and biofilm formation (S7 - S9 Figs). Competitive co-

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

infections with UTI89 and UTI89 *hsdS*^{MG1655} again revealed no competitive advantage for either strain, irrespective of the time-point or organ tested (Fig 4B).

To test whether UTI89 was unique in its ability to tolerate changes in methylation, we created a similar "methylation-switch" in MG1655 by inserting the UTI89 *hsdS* allele. We again found no differences between wt MG1655 and MG1655 *hsdS*^{UTI89} in the Biolog PM phenotype screen (S4 Table, S10 Fig). We therefore conclude that, at least in UTI89 and MG1655, substantial changes in DNA methylation across the genome are not merely tolerated but simply have no measurable effect on gene expression or laboratory-measured phenotypes.

Discussion

The EcoKI restriction modification system (RMS) was identified over a half century ago due to its role in the specific restriction of exogenous (phage) DNA (73). EcoKI is thus the prototypical example of a RMS, but subsequent studies also demonstrated that, particularly for *E. coli* but also for many other bacteria, this was an archetypal system as well. It was thus designated "Type I" in the RMS classification that followed. As more RMSs were discovered, they were initially assumed to play similar roles in host defense and horizontal gene transfer (81,82). In bacteria, one of the early indications that methylation could affect transcription arose from the study of regulation of RMSs themselves, some of which utilize methylation as a readout of expression level, effectively a type of product-inhibition feedback (83). The general observation that DNA methylation, particularly that mediated by orphan methyltransferases (such as Dam or CcrM, which may themselves have evolved from RMSs), could alter transcription (44,48,84), led to the discovery of a broad suite of associated roles in DNA metabolism, cell physiology, and virulence (12,51,55,85–87). An active research community continues to describe and characterize the additional roles that DNA methylation (from both restriction and orphan systems) play in bacteria (4,39,88), some of which can be very specific (67,69). We now show that, in contrast with expectations from

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

this literature, the archetypal EcoKI Type I RMS has nearly zero impact on gene regulation or any phenotype among more than 1000 growth conditions tested. These results generalize to two additional Type I RMSs from pathogenic strains of *E. coli*. We therefore conclude that these Type I RMSs have a strictly limited role in the classical defense against incoming foreign DNA, and the associated methylation of hundreds of adenines throughout the genome specifically avoids any consequential regulation of gene expression.

The paradigm of "one gene, one function" made the initial discovery that restriction systems could play a distinct epigenetic role in gene regulation somewhat surprising. Now, the idea that methylation can affect the interaction with DNA binding proteins is quite clear (12), and orphan methyltransferases were thought to have preserved DNA methylation for such non-defense (i.e. regulatory) functions (11). Perhaps the strongest evidence that gene regulation is a biologically selected function for RMSs comes from the discovery of phasevarions, in which phase variable expression (through recombination or simple sequence repeat variation) leads to rapid switching between RMSs with different methylation specificities, leading to differential regulation of genes that have particularly been shown to impact virulence (8). Type I systems are the least common of the four classes of RMSs, however, and generally have the longest recognition sequences (leading to fewer modified sites in a given genome) (11,39,59). Regardless, phasevarions composed of multiple Type I specificities have also been described, with similar effects on gene expression and virulence (35,36,89); in addition, non-phase variable Type I systems have also been shown to affect gene expression (67–69).

The three Type I RMSs studied here, EcoUTI, EcoCFTI, and EcoKI, each methylate many hundreds (427 to 754) of distinct sites in their respective host genomes (and similar predicted numbers in other *E. coli* genomes). We tested for phenotypic impacts using an *in vivo* murine model of urinary tract infection, a complex phenotype which demands that inoculated cells react to changing nutrient, fluid flow, and host immune conditions. In addition, *E. coli* are known to occupy different niches within the urinary tract, ranging from

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bladder epithelial cells (both within the cytoplasm and on the surface) to the lumen of the bladder and the parenchyma of the kidney (90). While we did not investigate these different niches individually, there was no overall difference in infection titers, even during competition experiments with a wild type strain, a strong indication that no virulence defect is present when EcoUTI is deleted. Complementing this complex phenotype, we used the Biolog PM platform to screen a "simple" growth phenotype in 1190 conditions, again finding no difference regardless of the presence or absence of the three Type I RMSs. Finally, we used RNA-seq to explore whether gene expression (either on a global scale or at the level of individual genes) was altered by changing methylation, again finding no differences attributable to DNA methylation. It remains possible that under some specific conditions, these Type I systems would indeed alter gene expression or phenotypes; indeed, methyltransferases such as *Escherichia coli* Dcm and *Mycobacterium tuberculosis* MamA alter gene expression primarily under specific growth conditions such as stationary phase and hypoxia, respectively (61,78,91). However, it seems that finding such a condition might serve as the "exception that proves the rule", given our data using both simple and complex phenotypes, across >1000 growth conditions, and measuring genome-wide gene expression under lab conditions. Notably, in the course of making one of the Type I RMS knockouts, we recovered a clone carrying a nonsynonymous mutation in a nucleotide biosynthesis operon; control experiments using this clone demonstrated that we could indeed detect differences in *in vivo* infections, PM arrays, and gene expression (manuscript in preparation).

Based on our results, we propose the concept of "regulation avoidance". This term describes a state where methylation (i.e. changing from unmethylated to methylated or vice versa) has no measurable impact on gene expression or any phenotype. We have shown that the Type I RMSs in UTI89, MG1655, and CFT073 all possess the property of regulation avoidance in their respective strains; furthermore, the MG1655 EcoKI and CFT073 EcoCFTI RMSs have regulation avoidance when introduced into UTI89, and the UTI89 EcoUTI RMS has regulation avoidance when introduced into MG1655. We strongly suspect that regulation

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

avoidance is not limited to the specific *E. coli* strains and Type I systems studied here, despite the fact that many RMSs have been shown to not have this property (65,67–69). Therefore, future studies will be required to characterize the regulatory role of other RMSs, potentially leading to general principles that may help explain when and why some RMSs have regulation avoidance (possibly only in some strains). At this point, we suspect that Type I systems, due to their longer recognition sequence and overall fewer methylated sites, will be more likely to be regulation avoidant than other RMSs.

A few additional implications of regulation avoidance merit discussion here. First, regulation avoidance likely is related to the fact that RMSs are mobile, or that they switch regularly, on an evolutionary time scale. A permanent and fixed pattern of methylation would likely result in adaptation of the host genome and host proteins to accommodate those sites where methylation overlaps binding sites for transcription factors or RNA polymerase. This would be similar to the process of amelioration of codon usage or compensatory mutations to minimize fitness defects from new plasmids, for example (92,93). Our data, showing that removal and replacement of each of three separate Type I systems has no detectable impact on regulation in UTI89, instead implies that any genomic adaptation would have occurred towards *both* methylated and unmethylated motifs. Second, some *E. coli* strains indeed do not have a Type I RMS, but the majority do; furthermore, RMSs are thought to be easily transferred between strains (also on an evolutionary time scale) (74,75). In particular, details of the structure of Type I RMS enable rapid switching of specificity, for example by evolution of or limited recombination of the target recognition domain of the *hsdS* gene (19,65). Our data on strains in which we installed two different Type I RMSs into UTI89 is consistent with this notion, and we speculate that the lack of any effect in deleting EcoUTI is actually driven by the presence of other Type I RMSs in the evolutionary history of UTI89. Third, given the presumed ease of evolving methylation specificity through mutation of the *hsdS* gene, regulation avoidance is possibly also mediated by adaptation of the recognition sequence of Type I RMSs. Finally, early examples of methylation regulating gene expression

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in bacteria came from studies of the regulation of expression of the RMSs themselves. One consistent theme from such studies is that some details of the regulation can be rationalized using a scenario where a new RMS is introduced into a bacterium: methylation is expressed first, while restriction is expressed only with a delay, thus preventing immediate degradation of the chromosome (94). This is similar to the requirements expected for introduction of a new toxin-antitoxin system (requiring antitoxin expression before toxin expression) (95), and further suggests that the introduction of a new RMS may be a regular occurrence. Of note, we have emphasized the evolutionary time scale here to contrast with phasevarions, which switch RMS specificity on a time scale much shorter than that required for genomic adaptation.

In summary, we have conclusively demonstrated that, for 3 different strains of *E. coli*, the Type I RMSs have very nearly zero effect on gene expression and no effect on any phenotype tested besides their originally described function of defense against incoming foreign DNA. We introduce the concept of “regulation avoidance” to highlight the contrast of our results with the weight of the published literature on non-defense functions of other RMSs and orphan methyltransferases. Regulation avoidance, in turn, leads to several insights into the evolutionary dynamics of RMSs, which suggest fruitful avenues for future study. In particular, how general is regulation avoidance among *E. coli*, and is it limited to Type I RMSs? Do other bacteria have regulation avoidant RMSs? Is regulation avoidance necessarily dependent upon preservation of DNA restriction function, or is evolutionary plasticity and exchange of methylation specificity sufficient? How much is regulation avoidance mediated by adaptation of DNA binding proteins, placement and/or context of binding sites in the genome, or modification of RMS specificity? Future studies addressing these questions may provide further insights into the impact that RMSs (and, by implication, phages and other mechanisms that introduce foreign DNA) have had on bacterial genome evolution.

Material and methods

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ethics statement

All animal experiments were approved by and performed in strict accordance with the Institutional Animal Care and Use Committee (IACUC) of A*STAR (Protocols #110605 and #130853).

Media and culture conditions

All strains were propagated in Lysogeny Broth (LB) at 37°C unless otherwise noted. Media was supplemented with ampicillin (100 µg/ml), kanamycin (50 µg/ml), or chloramphenicol (20 µg/ml) where required (all antibiotics from Sigma, Singapore). M9 minimal medium (1X M9 salts, 2mM Magnesium sulphate, 0.1mM Calcium chloride, 0.2% glucose) and Yeast extract-Casamino acids (YESCA) medium (10g/L Casamino acids, 1g/L Yeast extract) were used for specific experiments.

Strain and plasmid generation

Deletion mutants were generated using a λ-Red recombinase mediated strategy optimised for clinical strains of *E. coli*, with minor modifications (96). Briefly, a positive selection cassette encoding either kanamycin (*neo*) or chloramphenicol (*cat*) resistance was amplified from plasmid pKD4 or pKD3, respectively (97). Primers incorporated, at their 5'-end, 50bp of homology to the genomic locus being knocked out. The resultant PCR product was transformed into cells expressing λ-Red recombinase from vector pKM208, recovered at 37°C for 2 hours with shaking followed by 2 hours without shaking, then plated onto the appropriate antibiotic at 37°C overnight.

Strains containing a seamless marker-free replacement of a gene with a different allele were generated using a previously described negative selection strategy (98). Two successive rounds of λ-red recombinase mediated homologous recombination were performed, using the protocol described above, to first replace the wild type allele with a dual positive-negative

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

selection cassette (amplified from plasmids pSLC217 or pSLC246 instead of pKD4 or pKD3). A second round of recombination was then performed using a PCR product with the desired allele instead of a resistance cassette, and the second selection done on M9 supplemented with 0.2% rhamnose. Correct clones were confirmed by PCR and by sequencing the recombination junctions by Sanger sequencing (1st Base, Singapore).

The Type I RMS recognition motifs for UTI89 (5'-GAAG(N_7)TGG-3'), MG1655 (5'-AAC(N_6)GTGC-3'), and CFT073 (5'-GAG(N_7)GTCA-3') were incorporated into primers designed using an inverse PCR strategy to amplify plasmid pACYC184 (which does not natively possess any of these Type I recognition motifs). Linear plasmid amplicons were then digested using NotI, recircularized with T4 DNA ligase, then transformed into *E. coli* TOP10. When two copies of the Type I motifs were inserted, they were placed 1500bp apart (pACYC184 coordinates 381 and 1974). All strains, plasmids and primers (Sigma, Singapore) used in this study are listed in S6 and S7 Tables, respectively.

Transformation efficiency assay

Competent cells were prepared using log phase cultures obtained by sub-culturing overnight bacterial culture 1:100 in LB and growing at 37°C up to OD_{600} = 0.4 – 0.5. Cells were washed twice with sterile water followed by a final wash with sterile 10% glycerol and resuspension in 1/100 of the original culture volume of 10% glycerol; these were then stored at -80°C in 50 µl aliquots. Plasmids with 1 or 2 copies of the bipartite Type I motif were extracted from the appropriate wild type or Type I RMS mutants to obtain methylated or unmethylated preparations. Plasmids were extracted using the Hybrid-Q plasmid miniprep kit (GeneAll, South Korea) according to the manufacturer's recommended protocol. Competent cells were transformed with 100ng of each plasmid using 1mm electroporation cuvettes in a GenePulser XCELL system, at 400 Ω resistance, 25µF capacitance, and 1700V output voltage (Bio-Rad, Singapore). Cells were recovered in 1 ml of prewarmed LB at 37°C for 1 hour with shaking and plated on selective (chloramphenicol) and non-selective (LB) plates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Transformation efficiency was calculated by dividing the cfu/ml obtained on selective by that on non-selective plates per unit amount of plasmid DNA.

Mouse infections

In vivo infections were performed using a murine transurethral model of urinary tract infection (66). Briefly, bacterial strains were grown in Type I pili-inducing conditions by two passages in LB broth at 37°C for 24 hours without shaking; a 1:1000 dilution was made from the first to the second passage. Cells were then harvested by centrifugation and resuspended in sterile cold PBS to OD_{600} = 1. Type I piliation for each strain was evaluated by a hemagglutination assay and Type I phase assay as described previously (99). A 1:2 dilution of the PBS suspension (final OD_{600} = 0.5) was then used as the inoculum. 7-8 week old female C3H/HeN mice (InVivos, Singapore) were anaesthetized using isoflurane and 50 µl of inoculum (OD_{600} = 0.5, ~1-2 x 10^7 cfu/50 µl) was transurethrally instilled into the bladder using a syringe fitted with a 30 gauge needle covered with a polyethylene catheter (Product #427401, Thermo fisher scientific, USA). At specified times, mice were sacrificed and bladders and kidneys were harvested aseptically and homogenized in 1 ml and 0.8 ml of sterile PBS, respectively. Ten-fold serial dilutions were plated on appropriate selective plates to quantify bacterial loads. For co-infections, the inoculum consisted of a 1:1 mixture of two strains with an antibiotic resistance cassette inserted at the phage HK022 attachment site *attP* (100), each at 1-2 x 10^7 CFU/50 µl; otherwise, the procedure was identical to that described for the single infections above. To account for potential bias due to selection markers, co-infections were performed with an equal number of mice infected with strains with their selection marker combinations reversed. For example, co-infections comparing wild type (Kan^R) and mutant ($Chlor^R$) were also performed with wild type ($Chlor^R$) and mutant (Kan^R). Bacterial titres from each organ and starting inoculum were used to calculate the competitive index (CI) as follows: CI = (Output wild type / Output mutant) / (Input wild type / Input mutant).

RNA sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Stationary phase cells were obtained by propagating strains statically for two serial 24 hour passages at 37°C, with a 1:1000 dilution between passages (identical to the inoculum used for mouse infections). Log phase cells were obtained by diluting overnight cultures 1:100 in LB and growing to log phase (OD_{600} = 0.4 – 0.5) at 37°C with shaking. RNA was extracted with the RNeasy Mini kit (Qiagen, Singapore) from three biological replicates for both log and stationary phase samples using 7 x 10^8 cells. RNA quality was assessed using the Agilent RNA 6000 pico kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA); only replicates where all samples had an RNA integrity number (RIN) >= 7 were used. Ribosomal RNA (rRNA) depletion was done with the Ribo-Zero rRNA removal kit (Epicenter, USA) according to the manufacturer's recommended protocol. Libraries were generated using the ScriptSeq v2 RNA-seq library preparation kit (Epicenter, USA) according to the manufacturer's recommended protocol. Each uniquely indexed strand specific library was assessed for library size and amount using the Agilent DNA 1000 kit (Agilent Technologies, USA). After normalization and pooling, samples were sequenced on either the Illumina HiSeq 4000 or NextSeq sequencer with 2 x 151 bp or 2 x 76 bp reads (Illumina, USA).

Raw sequencing reads were mapped to their respective reference genomes: RefSeq accession GCF_000013265.1 for *E. coli* UTI89, GCF_000005845.2 for *E. coli* MG1655 and GCF_000007445.1 for *E. coli* CFT073 using BWA-MEM (version 0.7.10) with default parameters (101). HTseq was used to quantify sequencing reads mapping to predicted open reading frames (ORFs) (102). Ribosomal RNA (rRNA) and transfer RNA (tRNA) sequences (based on the corresponding Genbank RefSeq annotation) were filtered out of the data set. R (version 2.15.1) was used for differential expression analysis, using the edgeR package (103). Briefly, samples were normalized by TMM (trimmed median of means), common and tagwise dispersion factors were estimated using a negative binomial model, and then fold change values were calculated from these normalized counts. A false discovery rate (FDR) cutoff of <= 0.05 and a $\log_2$ fold change >= 1.5 were applied, resulting in the final set of differentially expressed genes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pacific Biosciences single molecule real time (SMRT) sequencing

Genomic DNA was extracted from log phase bacterial cultures grown in LB at 37°C with shaking and quantified using a Qubit 2.0 Fluorometer (Life Technologies, USA) using the dsDNA HS kit. 5 µg of DNA was sheared to 10 kbp using a g-Tube (Covaris, USA), and a SMRTbell library was made with the SMRTbell template prep kit 1.0 (Pacific Biosciences, USA) according to the manufacturer's instructions. Library quality and quantity were assessed using the Agilent DNA 12000 kit (Agilent Technologies, USA) and sequenced on the PacBio RS II sequencer (Pacific Biosciences, USA). Sequencing was performed using a single SMRTCell with P4-C2 enzyme chemistry using a 180 min movie. Reads were mapped back to the corresponding reference genome (as indicated under "RNA Sequencing") and methylated motifs were identified using the "RS_Modification_and_motif_analysis" algorithm in SMRT Analysis suite v2.3 using default parameters. Bases with a coverage of at least 25X and a kinetic score of at least 60 (default values for the RS_Modification_and_motif_analysis protocol) were identified as being methylated.

Quantitative RT-PCR

Samples were prepared as described under "RNA sequencing", using the RNeasy Mini kit (Qiagen, Singapore). 1 µg of RNA was treated with DNase I, RNase-free (Thermo fisher scientific, USA) at 37°C for 1 hour to remove any residual genomic DNA. Next, 500 ng of RNA was used for cDNA synthesis with SuperScript II Reverse Transcriptase and Random hexamers (Invitrogen, USA) according to the manufacturer's recommended protocol. Amplified cDNA was diluted 1:4 for all target genes and 1:400 for the *rrsA* internal control. Real-time PCR was performed with the KAPA SYBR FAST qPCR kit (KAPA Biosystems, USA) on a LightCycler 480 instrument (Roche, Singapore) with the following cycle: 95°C for 5 mins, followed by 40 cycles of 95°C for 30 seconds and 60°C for 30 seconds. Target-specific primers for qRT-PCR were designed using the IDT PrimerQuest tool (IDT, Singapore) and are listed in S7 Table. Relative fold change for target genes was calculated by the ΔΔCT method utilizing

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the 16S ribosomal gene *rrsA* as the internal control. Reverse transcriptase and negative controls were included in each run.

Motility Assay

Motility assays were performed as described previously with slight modifications (104). Strains were grown to log phase (OD_{600} = 0.4 – 0.5) by sub-culturing overnight bacterial cultures 1:100 in LB at 37°C. Bacteria were harvested by centrifugation (3200 g, 10 min) and resuspended in sterile PBS to OD_{600} = 0.4. Sterile 0.25% LB agar plates were prepared and stabbed once with each strain using sterile toothpicks. The soft agar plate was then incubated at 37°C for 7 – 8 hours. Motility was calculated by measuring the diameter of the bacterial motile front. Distances were normalized to a wild type control included in each experiment.

Biofilm Assay

A 96-well Crystal Violet biofilm assay was performed as previously described (105). Briefly, strains were grown to log phase (OD_{600} = 0.4 – 0.5) by sub-culturing overnight bacterial culture 1:100 in LB at 37°C and resuspended in sterile PBS to OD_{600} = 0.4. 96-well clear flat bottom poly vinyl chloride (PVC) plates (Product #01816049, Costar, Singapore) were seeded with 200 µl of sterile media (LB or YESCA) and 5 µl of the PBS suspension. PVC plates were incubated for the specified times at 26°C or 37°C in a humidified chamber. Plates were washed once with water and stained with Crystal Violet for 30 mins. Excess stain was removed by washing thrice with water. Residual stain was then dissolved in 200 µl of 50% ethanol, with care taken to avoid disturbing the biofilm. The amount of biofilm was quantified by measuring OD at 590nm using a Sunrise 96 well microplate absorbance reader (Tecan, Switzerland). The biofilm produced by each test strain was normalized to that of the corresponding wild type strain.

Growth Curves

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Growth curves for bacterial strains were measured using the Bioscreen C instrument (Bioscreen, Finland). Strains were grown to log phase (OD_{600} = 0.4 – 0.5) by sub-culturing overnight bacterial culture 1:100 in LB at 37°C and resuspended in sterile PBS to OD_{600} = 0.4. 5 µl of this normalized bacterial suspension was then inoculated into 145 µl of the desired growth media (LB (rich) or M9 (minimal)) in triplicates. Plates were incubated at 37°C and OD_{600} was measured every 15 mins for 20 hours.

Biolog Phenotype Microarray (PM)

PM assays (72) were performed using PM plates 1 to 20 by the PM services group at Biolog, USA using their standard protocol for *E. coli*. Succinate was added as the carbon source for PM plates 3-20 and all experiments with UTI89 (a niacin auxotroph) included 1 µg/ml niacin. All experiments were done with biological duplicates. Briefly, PM plates contain: PM1-2 Carbon sources, PM3 Nitrogen sources, PM4 Phosphorous and Sulphur sources, PM5 Nutrient supplements, PM6-8 Peptide nitrogen sources, PM9 Osmolytes, PM10 pH, and PM11-20 Inhibitors for different metabolic pathways. The PM services group analyzed growth curves, as measured colorimetrically, using their own proprietary software. Gain or loss of a phenotype/resistance was called by the PM services group, again using their proprietary software, which identified phenotypic differences based on the presence of a height difference between the strains in both replicates and a quality score >150 as cutoffs.

Figures

Fig 1. UTI89 Type I RMS EcoUTI functions only as a restriction system, with no apparent regulatory role. (A) Transformation efficiency assay using plasmids bearing 0, 1, or 2 copies (as indicated on the x-axis) of the UTI89 Type I RMS motif (5'-CCA(N_7)CTTC-3'). Recipient cells were wild type UTI89 and the isogenic *ΔhsdSMR* mutant, as indicated by the labels below the x-axis. Unmethylated and methylated plasmid preparations were used to transform each strain, as indicated by the legend at the top right. An unpaired t-test was used to identify significant differences between plasmids with 0, 1 and 2 methylation sites for

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

595 both preparations and strains; * p <0.05, n = 3 biological replicates. Data represents mean ±
596 standard deviation (s.d.) of log transformed values. **(B)** Competitive index (CI) for in vivo co-
597 infections with wt UTI89 and UTI89*ΔhsdSMR*. Bladder and kidney pairs (as indicated on the
598 x-axis) were aseptically harvested at 1 or 7 days post infection (dpi), as indicated by the
599 labels below the x-axis, and plated on appropriate selective plates for calculation of CI. A
600 Wilcoxon signed rank test was used to test for a significant difference of log CI from 0; * p
601 <0.05, n = 20 mice/time point, performed as 2 biological replicates. Each point represents
602 data from a single mouse; horizontal lines represent the median. **(C)** RNA sequencing
603 comparing logarithmic phase transcriptomes of UTI89 and UTI89*ΔhsdSMR.* Left, a volcano
604 plot of log FDR against log fold change. Right, qq-plots showing the distribution of
605 uncorrected p-values. Significantly differentially expressed genes ($\log_2$ fold change >1.5 and
606 $\log_{10}$ FDR <0.05) are labelled and colored either red (deleted genes/polar effects) or blue
607 (validated as false positive by qRT-PCR). n = 3 biological replicates. **(D)** Phenotype
608 microarray (PM) panel with plates PM1 to 20 comparing wt UTI89 and UTI89*ΔhsdSMR*.
609 Each plate is represented as a 12x8 grid of growth curves (red (wt), green (*ΔhsdSMR*), and
610 yellow (overlap) on a gray background). Each growth curve plots growth (measured
611 colorimetrically) (y-axis) against time (x-axis). Wells representing conditions where a height
612 difference was observed between the strains in both replicates are boxed in black, and wells
613 which also have a quality score >150 are considered significant and boxed in red. n = 2
614 biological replicates.

615 **Fig 2. Type I RMSs from *E. coli* CFT073 and MG1655 do not affect gene expression.**
616 RNA sequencing comparing logarithmic phase transcriptomes of **(A)** MG1655 and
617 MG1655*ΔhsdSMR*, and **(B)** CFT073 and CFT073*ΔhsdSMR.* Left, a volcano plot of log FDR
618 against log fold change. Right, qq-plots showing the distribution of uncorrected p-values.
619 Significantly differentially expressed genes ($\log_2$ fold change >1.5 and $\log_{10}$ FDR <0.05) are
620 labelled and colored either red (deleted genes/polar effects), brown (differentially expressed

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

621 genes) or blue (validated as false positive by qRT-PCR). n = 2 (**A** MG1655) or 3 (**B** CFT073)
622 biological replicates.

623 **Fig 3. Installation of different Type I methylation specificities into UTI89. (A)** Graphical
624 representation of the host specificity determinant (*hsd*) locus showing the amino acid identity
625 between Type I RMSs EcoKI, EcoUTI, and EcoCFTI from *E. coli* strains MG1655, UTI89,
626 and CFT073 respectively. The subclassification for each system is indicated on the left, with
627 the total length of DNA spanned by the *hsd* genes shown in parentheses. The *hsd* genes are
628 shown in different colors for visualization and drawn to scale, with transcription direction
629 indicated by the arrow. Dotted lines further connect the corresponding beginnings and ends
630 of each *hsd* gene in different systems. Genes adjacent to the *hsd* genes are shown in gray.
631 The percentage identity between *hsd* genes from different systems is indicated; green >99%
632 and black <50%. **(B)** Transformation efficiency assay using plasmids bearing 0, 1, or 2
633 copies (as indicated on the x-axis) of the MG1655 (5'-AAC(N_6)GTGC-3'), CFT073 (5'-
634 GAG(N_7)GTCA-3') and UTI89 (5'-CCA(N_7)CTTC-3') Type I RMS motifs. Recipient cells were
635 UTI89 $hsdS^{MG1655}$, UTI89 $hsdSMR^{CFT073}$ and UTI89 $hsdS^{UTI89}$, as indicated by the labels
636 below the x-axis. Unmethylated and methylated plasmid preparations were used to
637 transform each strain, as indicated by the legend at the top right. An unpaired t-test was
638 used to identify significant differences between plasmids with 0, 1 and 2 methylation sites for
639 both preparations and strains; * $p < 0.05$, n = 3 biological replicates. Data represents mean ±
640 s.d. of log transformed values.

641 **Fig 4. Heterologous Type I methylation in UTI89 does not affect virulence and gene**
642 **expression. (A)** Distribution of heterologous Type I methylation sites in the UTI89 genome.
643 Methylation sites for different systems are indicated by tick marks on different concentric
644 circles; from the outer to the inner circle, UTI89 (747 sites), MG1655 (628 sites), and
645 CFT073 (415 sites) sites are represented. Positions on the positive and negative strands are
646 represented by colored tick marks as indicated in the legend at the bottom right, marked in
647 red and blue respectively. **(B)** Competitive index (CI) for in vivo co-infections with wt UTI89

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and UTI89 *hsdS*^{MG1655}. Bladder and kidney pairs (as indicated on the x-axis) were aseptically harvested at 1 or 7 days post infection (dpi), as indicated by the labels below the x-axis, and plated on appropriate selective plates for calculation of CI. A Wilcoxon signed rank test was used to test for a significant difference of log CI from 0; * p <0.05, n = 25 mice (1dpi) and 20 mice (7dpi), performed as 2 biological replicates. Each point represents data from a single mouse; horizontal lines represent the median. **(C and D)** RNA sequencing comparing logarithmic phase transcriptomes of **C)** UTI89 and UTI89 *hsdS*^{MG1655}, and **D)** UTI89 and UTI89 *hsdSMR*^{CFT073}. Left, a volcano plot of log FDR against log fold change. Right, qq-plots showing the distribution of uncorrected p-values. Significantly differentially expressed genes ($\log_2$ fold change >1.5 and $\log_{10}$ FDR <0.05) are labelled and colored either red (deleted genes/polar effects) or brown (differentially expressed genes). n = 3 biological replicates.

Acknowledgements

The authors thank members of the Chen lab, Shu Sin Chng, William Burkholder, and Grishma Rane for helpful discussions and comments on the manuscript. The authors thank Steven Turner, Jonas Korlach, and Tyson Clark from Pacific Biosciences and Jacqueline L.Y. Chee for assisting with analysis of UTI89 methylation specificities using SMRT sequencing. The authors would also like to thank Adriano Baćac for additional technical assistance. Finally, the authors wish to thank the Next Generation Sequencing Platform at the Genome Institute of Singapore for technical help and guidance designing sequencing experiments.

Authors' contributions

KSM performed the experiments. KSM and SLC designed the experiments, analyzed the data, and wrote the manuscript. Both authors read and approved the final manuscript.

Availability of data and materials

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

672 The PacBio sequencing data is available in the GenBank Short Read Archive under
673 BioProject PRJNA474982 (106). The RNA-seq datasets generated and analysed during the
674 current study are available in the GenBank Short Read Archive under BioProject
675 PRJNA675406.

676 **Funding**

677 This work was supported by the National Research Foundation, Singapore (NRF-RF2010-
678 10), the Singapore Ministry of Health's National Medical Research Council under two
679 Clinician-Scientist Individual Research Grants (NMRC/CIRG/1467/2017 and
680 NMRC/CIRG/1358/2013), and the Genome Institute of Singapore (GIS)/Agency for Science,
681 Technology and Research (A*STAR). The funders had no role in the design, collection,
682 analysis, or interpretation of the data. The funders had no role in the writing of the
683 manuscript.

684 **References**

685 1. Holliday R. Epigenetics: A Historical Overview [Internet]. Vol. 1, Epigenetics. 2006. p.
686 76–80. Available from: http://dx.doi.org/10.4161/epi.1.2.2762

687 2. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond.
688 Nat Rev Genet. 2012;13(7):484–92.

689 3. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications
690 Regulate Gene Expression [Internet]. Vol. 32, Trends in Genetics. 2016. p. 42–56.
691 Available from: http://dx.doi.org/10.1016/j.tig.2015.10.007

692 4. Sánchez-Romero MA, Casadesús J. The bacterial epigenome. Nat Rev Microbiol. 2020
693 Jan;18(1):7–20.

694 5. Shorter J, Lindquist S. Prions as adaptive conduits of memory and inheritance. Nat Rev
695 Genet. 2005 Jun;6(6):435–50.

696 6. Veening JW, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in
697 bacteria. Annu Rev Microbiol. 2008;62:193–210.

698 7. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev
699 Genet. 2013 Mar;14(3):204–20.

700 8. Seib KL, Srikhanta YN, Atack JM, Jennings MP. Epigenetic Regulation of Virulence and
701 Immunoevasion by Phase-Variable Restriction-Modification Systems in Bacterial
702 Pathogens. Annu Rev Microbiol. 2020 Sep 8;74:655–71.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

703 9. Tan A, Atack JM, Jennings MP, Seib KL. The Capricious Nature of Bacterial Pathogens:
704 Phasevarions and Vaccine Development. Front Immunol. 2016;7:586.

705 10. Iyer LM, Zhang D, Aravind L. Adenine methylation in eukaryotes: Apprehending the
706 complex evolutionary history and functional potential of an epigenetic modification.
707 Bioessays. 2016 Jan;38(1):27–40.

708 11. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, et al. The
709 Epigenomic Landscape of Prokaryotes. PLoS Genet. 2016 Feb;12(2):e1005854.

710 12. Wion D, Casadesus J. N6-methyl-adenine: an epigenetic signal for DNA-protein
711 interactions. Nat Rev Microbiol. 2006 Mar;4(3):183–92.

712 13. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian
713 development and disease. Nat Rev Mol Cell Biol. 2019 Oct;20(10):590–607.

714 14. Skvortsova K, Iovino N, Bogdanović O. Functions and mechanisms of epigenetic
715 inheritance in animals. Nat Rev Mol Cell Biol. 2018 Dec;19(12):774–90.

716 15. Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of
717 Bertani and Weigle). Microbiol Mol Biol Rev. 2000 Jun;64(2):412–34.

718 16. Davis BM, Chao MC, Waldor MK. Entering the era of bacterial epigenomics with single
719 molecule real time DNA sequencing. Curr Opin Microbiol. 2013 Apr;16(2):192–8.

720 17. Pingoud A, Wilson GG, Wende W. Type II restriction endonucleases--a historical
721 perspective and more. Nucleic Acids Res. 2014 Jul;42(12):7489–527.

722 18. Rao DN, Dryden DT, Bheemanaik S. Type III restriction-modification enzymes: a
723 historical perspective. Nucleic Acids Res. 2014 Jan;42(1):45–55.

724 19. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. Type I restriction enzymes and their
725 relatives. Nucleic Acids Res. 2014 Jan;42(1):20–44.

726 20. Loenen WA, Raleigh EA. The other face of restriction: modification-dependent enzymes.
727 Nucleic Acids Res. 2014 Jan;42(1):56–69.

728 21. Srikhanta YN, Fox KL, Jennings MP. The phasevarion: phase variation of type III DNA
729 methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol.
730 2010 Mar;8(3):196–206.

731 22. Atack JM, Guo C, Yang L, Zhou Y, Jennings MP. DNA sequence repeats identify
732 numerous Type I restriction-modification systems that are potential epigenetic regulators
733 controlling phase-variable regulons; phasevarions. FASEB J. 2020 Jan;34(1):1038–51.

734 23. Atack JM, Yang Y, Seib KL, Zhou Y, Jennings MP. A survey of Type III restriction-
735 modification systems reveals numerous, novel epigenetic regulators controlling phase-
736 variable regulons; phasevarions. Nucleic Acids Res. 2018 Apr 20;46(7):3532–42.

737 24. Atack JM, Guo C, Litfin T, Yang L, Blackall PJ, Zhou Y, et al. Systematic Analysis of
738 REBASE Identifies Numerous Type I Restriction-Modification Systems with Duplicated,
739 Distinct hsdS Specificity Genes That Can Switch System Specificity by Recombination.
740 mSystems [Internet]. 2020 Jul 28;5(4). Available from:
741 http://dx.doi.org/10.1128/mSystems.00497-20

742 25. Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, Kwok T, Grimmond SM, et al.
743 Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. PLoS One.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

744 2011;6(12):e27569.

745 26. Srikhanta YN, Gorrell RJ, Power PM, Tsyganov K, Boitano M, Clark TA, et al.
746 Methylomic and phenotypic analysis of the ModH5 phasevarion of Helicobacter pylori.
747 Sci Rep. 2017;7(1):16140.

748 27. Tan A, Hill DM, Harrison OB, Srikhanta YN, Jennings MP, Maiden MC, et al. Distribution
749 of the type III DNA methyltransferases modA, modB and modD among Neisseria
750 meningitidis genotypes: implications for gene regulation and virulence. Sci Rep.
751 2016;6:21015.

752 28. Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, Harrison OB, et al.
753 Phasevarions mediate random switching of gene expression in pathogenic Neisseria.
754 PLoS Pathog. 2009 Apr;5(4):e1000400.

755 29. Kwiatek A, Mrozek A, Bacal P, Piekarowicz A, Adamczyk-Poplawska M. Type III
756 Methyltransferase M.NgoAX from Neisseria gonorrhoeae FA1090 Regulates Biofilm
757 Formation and Interactions with Human Cells. Front Microbiol. 2015;6:1426.

758 30. Atack JM, Srikhanta YN, Fox KL, Jurcisek JA, Brockman KL, Clark TA, et al. A biphasic
759 epigenetic switch controls immunoevasion, virulence and niche adaptation in non-
760 typeable Haemophilus influenzae. Nat Commun. 2015;6:7828.

761 31. Brockman KL, Jurcisek JA, Atack JM, Srikhanta YN, Jennings MP, Bakaletz LO. ModA2
762 Phasevarion Switching in Nontypeable Haemophilus influenzae Increases the Severity
763 of Experimental Otitis Media. J Infect Dis. 2016;214(5):817–24.

764 32. Blakeway LV, Power PM, Jen FE, Worboys SR, Boitano M, Clark TA, et al. ModM DNA
765 methyltransferase methylome analysis reveals a potential role for Moraxella catarrhalis
766 phasevarions in otitis media. FASEB J. 2014 Dec;28(12):5197–207.

767 33. Blakeway LV, Tan A, Jurcisek JA, Bakaletz LO, Atack JM, Peak IR, et al. The Moraxella
768 catarrhalis phase-variable DNA methyltransferase ModM3 is an epigenetic regulator
769 that affects bacterial survival in an in vivo model of otitis media. BMC Microbiol. 2019
770 Dec 9;19(1):276.

771 34. Srikhanta YN, Fung KY, Pollock GL, Bennett-Wood V, Howden BP, Hartland EL.
772 Phasevarion-Regulated Virulence in the Emerging Pediatric Pathogen Kingella kingae.
773 Infect Immun [Internet]. 2017 Dec;85(12). Available from:
774 http://dx.doi.org/10.1128/IAI.00319-17

775 35. Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, et al. A random six-
776 phase switch regulates pneumococcal virulence via global epigenetic changes. Nat
777 Commun. 2014;5:5055.

778 36. Li J, Li JW, Feng Z, Wang J, An H, Liu Y, et al. Epigenetic Switch Driven by DNA
779 Inversions Dictates Phase Variation in Streptococcus pneumoniae. PLoS Pathog. 2016
780 Jul;12(7):e1005762.

781 37. Oliver MB, Basu Roy A, Kumar R, Lefkowitz EJ, Swords WE. Streptococcus
782 pneumoniae TIGR4 Phase-Locked Opacity Variants Differ in Virulence Phenotypes.
783 mSphere [Internet]. 2017 Nov;2(6). Available from:
784 http://dx.doi.org/10.1128/mSphere.00386-17

785 38. Seshasayee AS, Singh P, Krishna S. Context-dependent conservation of DNA
786 methyltransferases in bacteria. Nucleic Acids Res. 2012 Aug;40(15):7066–73.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

787 39. Oliveira PH, Fang G. Conserved DNA Methyltransferases: A Window into Fundamental
788 Mechanisms of Epigenetic Regulation in Bacteria. Trends Microbiol [Internet]. 2020 May
789 13; Available from: http://dx.doi.org/10.1016/j.tim.2020.04.007

790 40. Julio SM, Heithoff DM, Provenzano D, Klose KE, Sinsheimer RL, Low DA, et al. DNA
791 adenine methylase is essential for viability and plays a role in the pathogenesis of
792 Yersinia pseudotuberculosis and Vibrio cholerae. Infect Immun. 2001 Dec;69(12):7610–
793 5.

794 41. Payelleville A, Lanois A, Gislard M, Dubois E, Roche D, Cruveiller S, et al. DNA
795 Adenine Methyltransferase (Dam) Overexpression Impairs Photorhabdus luminescens
796 Motility and Virulence. Front Microbiol. 2017;8:1671.

797 42. Falker S, Schmidt MA, Heusipp G. DNA methylation in Yersinia enterocolitica: role of
798 the DNA adenine methyltransferase in mismatch repair and regulation of virulence
799 factors. Microbiology. 2005 Jul;151(Pt 7):2291–9.

800 43. Marinus MG, Casadesus J. Roles of DNA adenine methylation in host–pathogen
801 interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev.
802 2009 May 1;33(3):488–503.

803 44. Oshima T, Wada C, Kawagoe Y, Ara T, Maeda M, Masuda Y, et al. Genome-wide
804 analysis of deoxyadenosine methyltransferase-mediated control of gene expression in
805 Escherichia coli. Mol Microbiol. 2002 Aug;45(3):673–95.

806 45. Chao MC, Zhu S, Kimura S, Davis BM, Schadt EE, Fang G, et al. A Cytosine
807 Methyltransferase Modulates the Cell Envelope Stress Response in the Cholera
808 Pathogen [corrected]. PLoS Genet. 2015 Nov;11(11):e1005666.

809 46. Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S. The core
810 genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in
811 orchestrating gene expression in Helicobacter pylori. Nucleic Acids Res. 2019 Mar
812 18;47(5):2336–48.

813 47. Kumar S, Karmakar BC, Nagarajan D, Mukhopadhyay AK, Morgan RD, Rao DN. N4-
814 cytosine DNA methylation regulates transcription and pathogenesis in Helicobacter
815 pylori. Nucleic Acids Res. 2018 Apr 20;46(7):3815.

816 48. Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J. The functions of DNA
817 methylation by CcrM in Caulobacter crescentus: a global approach. Nucleic Acids Res.
818 2014 Apr;42(6):3720–35.

819 49. Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, Sekulovic O, et al. Epigenomic
820 characterization of Clostridioides difficile finds a conserved DNA methyltransferase that
821 mediates sporulation and pathogenesis. Nature Microbiology. 2020 Jan 1;5(1):166–80.

822 50. Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ. An essential role for DNA adenine
823 methylation in bacterial virulence. Science. 1999;284(5416):967–70.

824 51. Heusipp G, Falker S, Schmidt MA. DNA adenine methylation and bacterial
825 pathogenesis. Int J Med Microbiol. 2007 Feb;297(1):1–7.

826 52. Erova TE, Pillai L, Fadl AA, Sha J, Wang S, Galindo CL, et al. DNA adenine
827 methyltransferase influences the virulence of Aeromonas hydrophila. Infect Immun.
828 2006 Jan;74(1):410–24.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

53. Pouillot F, Fayolle C, Carniel E. A putative DNA adenine methyltransferase is involved in Yersinia pseudotuberculosis pathogenicity. Microbiology. 2007 Aug;153(Pt 8):2426–34.

54. Mehling JS, Lavender H, Clegg S. A Dam methylation mutant of Klebsiella pneumoniae is partially attenuated. FEMS Microbiol Lett. 2007 Mar;268(2):187–93.

55. Stephenson SA, Brown PD. Epigenetic Influence of Dam Methylation on Gene Expression and Attachment in Uropathogenic Escherichia coli. Front Public Health. 2016;4:131.

56. Hernday A, Krabbe M, Braaten B, Low D. Self-perpetuating epigenetic pili switches in bacteria. Proc Natl Acad Sci U S A. 2002;99 Suppl 4:16470–6.

57. Wallecha A, Munster V, Correnti J, Chan T, van der Woude M. Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. J Bacteriol. 2002 Jun;184(12):3338–47.

58. Adhikari S, Curtis PD. DNA methyltransferases and epigenetic regulation in bacteria. FEMS Microbiol Rev. 2016 Sep;40(5):575–91.

59. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015 Jan;43(Database issue):D298–9.

60. Chen SL, Hung C-S, Xu J, Reigstad CS, Magrini V, Sabo A, et al. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. Proc Natl Acad Sci U S A. 2006 Apr 11;103(15):5977–82.

61. Militello KT, Simon RD, Qureshi M. Conservation of Dcm-mediated cytosine DNA methylation in Escherichia coli. FEMS microbiology [Internet]. 2012; Available from: https://academic.oup.com/femsle/article-abstract/328/1/78/505282

62. Kasarjian JK, Iida M, Ryu J. New restriction enzymes discovered from Escherichia coli clinical strains using a plasmid transformation method. Nucleic Acids Res. 2003;31(5):e22.

63. Studier FW, Bandyopadhyay PK. Model for how type I restriction enzymes select cleavage sites in DNA. Proc Natl Acad Sci U S A. 1988 Jul;85(13):4677–81.

64. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. 2003 Apr 1;31(7):1805–12.

65. Phillips ZN, Husna A-U, Jennings MP, Seib KL, Atack JM. Phasevarions of bacterial pathogens - phase-variable epigenetic regulators evolving from restriction-modification systems. Microbiology. 2019 Sep;165(9):917–28.

66. Hung CS, Dodson KW, Hultgren SJ. A murine model of urinary tract infection. Nat Protoc. 2009;4(8):1230–43.

67. Nye TM, Jacob KM, Holley EK, Nevarez JM, Dawid S, Simmons LA, et al. DNA methylation from a Type I restriction modification system influences gene expression and virulence in Streptococcus pyogenes. PLoS Pathog. 2019 Jun;15(6):e1007841.

68. Doberenz S, Eckweiler D, Reichert O, Jensen V, Bunk B, Sproer C, et al. Identification of a Pseudomonas aeruginosa PAO1 DNA Methyltransferase, Its Targets, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Physiological Roles. MBio [Internet]. 2017;8(1). Available from: http://dx.doi.org/10.1128/mBio.02312-16

69. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet. 2014 Apr;10(4):e1004272.

70. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, et al. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. Nat Biotechnol. 2012 Dec;30(12):1232–9.

71. Casselli T, Tourand Y, Scheidegger A, Arnold WK, Proulx A, Stevenson B, et al. DNA Methylation by Restriction Modification Systems Affects the Global Transcriptome Profile in Borrelia burgdorferi. J Bacteriol [Internet]. 2018 Dec 15;200(24). Available from: http://dx.doi.org/10.1128/JB.00395-18

72. Bochner BR. Global phenotypic characterization of bacteria. FEMS Microbiol Rev. 2009 Jan;33(1):191–205.

73. Loenen WA. Tracking EcoKI and DNA fifty years on: a golden story full of surprises. Nucleic Acids Res. 2003;31(24):7059–69.

74. Barcus VA, Titheradge AJ, Murray NE. The diversity of alleles at the hsd locus in natural populations of Escherichia coli. Genetics. 1995 Aug;140(4):1187–97.

75. Sibley MH, Raleigh EA. Cassette-like variation of restriction enzyme genes in Escherichia coli C and relatives. Nucleic Acids Res. 2004 Jan 16;32(2):522–34.

76. Titheradge AJ, King J, Ryu J, Murray NE. Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. Nucleic Acids Res. 2001;29(20):4195–205.

77. Nye TM, van Gijtenbeek LA, Stevens AG, Schroeder JW, Randall JR, Matthews LA, et al. Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in Bacillus subtilis. Nucleic Acids Res [Internet]. 2020 Apr 23; Available from: http://dx.doi.org/10.1093/nar/gkaa266

78. Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC, et al. DNA methylation impacts gene expression and ensures hypoxic survival of Mycobacterium tuberculosis. PLoS Pathog. 2013;9(7):e1003419.

79. Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, et al. Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. Proc Natl Acad Sci U S A. 2013;110(48):E4658–67.

80. Jen FE-C, Scott AL, Tan A, Seib KL, Jennings MP. Random Switching of the ModA11 Type III DNA Methyltransferase of Neisseria meningitidis Regulates Entner-Doudoroff Aldolase Expression by a Methylation Change in the eda Promoter Region. J Mol Biol [Internet]. 2020 Sep 4; Available from: http://dx.doi.org/10.1016/j.jmb.2020.08.024

81. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol. 2005 Sep;3(9):711–21.

82. Oliveira PH, Touchon M, Rocha EPC. Regulation of genetic flux between bacteria by restriction–modification systems. Proc Natl Acad Sci U S A. 2016 May 17;113(20):5658–63.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

913 83. Beletskaya IV, Zakharova MV, Shlyapnikov MG, Semenova LM, Solonin AS. DNA
914 methylation at the CfrBI site is involved in expression control in the CfrBI restriction-
915 modification system. Nucleic Acids Res. 2000 Oct 1;28(19):3817–22.

916 84. Lobner-Olesen A, Marinus MG, Hansen FG. Role of SeqA and Dam in Escherichia coli
917 gene expression: a global/microarray analysis. Proc Natl Acad Sci U S A.
918 2003;100(8):4672–7.

919 85. Cohen NR, Ross CA, Jain S, Shapiro RS, Gutierrez A, Belenky P, et al. A role for the
920 bacterial GATC methylome in antibiotic stress survival. Nat Genet. 2016
921 May;48(5):581–6.

922 86. Campellone KG, Roe AJ, Lobner-Olesen A, Murphy KC, Magoun L, Brady MJ, et al.
923 Increased adherence and actin pedestal formation by dam-deficient
924 enterohaemorrhagic Escherichia coli O157:H7. Mol Microbiol. 2007 Mar;63(5):1468–81.

925 87. Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. Microbiol Mol
926 Biol Rev. 2006 Sep;70(3):830–56.

927 88. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to
928 cellular defense. Microbiol Mol Biol Rev. 2013 Mar;77(1):53–72.

929 89. Ben-Assa N, Coyne MJ, Fomenkov A, Livny J, Robins WP, Muniesa M, et al. Analysis
930 of a phase-variable restriction modification system of the human gut symbiont
931 Bacteroides fragilis. Nucleic Acids Res. 2020 Nov 4;48(19):11040–53.

932 90. Terlizzi ME, Gribaudo G, Maffei ME. UroPathogenic Escherichia coli (UPEC) Infections:
933 Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial
934 Strategies. Front Microbiol. 2017 Aug 15;8:1566.

935 91. Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, et al. Genomics
936 of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase
937 transcription. Nat Commun. 2012;3:886.

938 92. Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. Parallel compensatory
939 evolution stabilizes plasmids across the parasitism-mutualism continuum. Curr Biol.
940 2015 Aug 3;25(15):2034–9.

941 93. López JL, Lozano MJ, Lagares A Jr, Fabre ML, Draghi WO, Del Papa MF, et al. Codon
942 Usage Heterogeneity in the Multipartite Prokaryote Genome: Selection-Based Coding
943 Bias Associated with Gene Location, Expression Level, and Ancestry. MBio [Internet].
944 2019 May 28;10(3). Available from: http://dx.doi.org/10.1128/mBio.00505-19

945 94. Morozova N, Sabantsev A, Bogdanova E, Fedorova Y, Maikova A, Vedyaykin A, et al.
946 Temporal dynamics of methyltransferase and restriction endonuclease accumulation in
947 individual cells after introducing a restriction-modification system. Nucleic Acids Res.
948 2016 Jan 29;44(2):790–800.

949 95. Fraikin N, Goormaghtigh F, Van Melderen L. Type II Toxin-Antitoxin Systems: Evolution
950 and Revolutions. J Bacteriol [Internet]. 2020 Mar 11;202(7). Available from:
951 http://dx.doi.org/10.1128/JB.00763-19

952 96. Murphy KC, Campellone KG. Lambda Red-mediated recombinogenic engineering of
953 enterohemorrhagic and enteropathogenic E. coli. BMC Mol Biol. 2003;4:11.

954 97. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640–5.

98. Khetrapal V, Mehershahi K, Rafee S, Chen S, Lim CL, Chen SL. A set of powerful negative selection systems for unmodified Enterobacteriaceae. Nucleic Acids Res. 2015;43(13):e83.

99. Chen SL, Hung CS, Pinkner JS, Walker JN, Cusumano CK, Li Z, et al. Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. Proc Natl Acad Sci U S A. 2009;106(52):22439–44.

100. Rossignol M, Moulin L, Boccard F. Phage HK022-based integrative vectors for the insertion of genes in the chromosome of multiply marked Escherichia coli strains. FEMS Microbiol Lett. 2002;213(1):45–9.

101. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

102. Anders S, Pyl PT, Huber W. HTSeq: Analysing high-throughput sequencing data with Python. 2010.

103. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

104. Wright KJ, Seed PC, Hultgren SJ. Uropathogenic Escherichia coli flagella aid in efficient urinary tract colonization. Infect Immun. 2005 Nov;73(11):7657–68.

105. Hadjifrangiskou M, Gu AP, Pinkner JS, Kostakioti M, Zhang EW, Greene SE, et al. Transposon mutagenesis identifies uropathogenic Escherichia coli biofilm factors. J Bacteriol. 2012 Nov;194(22):6195–205.

106. Fenlon SN, Chee YC, Chee JLY, Choy YH, Khng AJ, Liow LT, et al. Sequencing of E. coli strain UTI89 on multiple sequencing platforms. BMC Res Notes. 2020 Oct 20;13(1):487.

107. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010 Jun;7(6):461–5.

108. Clark TA, Lu X, Luong K, Dai Q, Boitano M, Turner SW, et al. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. BMC Biol. 2013;11:4.

Supporting information

S1 Table. Methylated motifs identified by SMRT sequencing in *E. coli* UTI89 and UTI89*ΔhsdSMR*. Modified positions are underlined. DNA modifications: N6-methyladenine (6mA) and C5-methylcytosine (5mC). N=A, T, G or C and W=A or T. ***** low percentage or

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

absence of 5mC bases is due to the weak signal-to-noise ratio for this modification, requiring higher coverage or chemical conversion prior to sequencing (107,108).

S2 Table. Logarithmic phase differentially expressed genes by RNA-seq. Genes with $\log_{10}$ FDR <=0.05 and $\log_2$ fold change =>1.5 are considered significant and highlighted in bold. *denotes deleted genes/alleles and polar effects.

S3 Table. Stationary phase differentially expressed genes by RNA-seq. Genes with $\log_{10}$ FDR <=0.05 and $\log_2$ fold change =>1.5 are considered significant and highlighted in bold. *denotes deleted genes/alleles and polar effects.

S4 Table. Phenotypic differences identified by PM due to altered *E. coli* Type I methylation. Type I methylation mutants are compared to corresponding wt strains and phenotypic differences listed in descending value of quality score. Phenotypes with quality score >150 are considered significant and highlighted in bold. * denotes resistance conferred by presence of kanamycin selection cassette at *hsdSMR* locus.

S5 Table. Distribution of Type I methylation motifs in *E. coli* UTI89, MG1655 and CFT073 genomes. *UTI89 genome includes the 5,065,741bp chromosome and 114,230bp plasmid (pUTI89).

S6 Table. List of strains and plasmids.

S7 Table. List of primers.

S1 Fig. Loss of native Type I methylation does not alter UTI89 urovirulence. Wild type UTI89 or isogenic UTI89*ΔhsdSMR* strains were transurethrally inoculated (2×10^7 cfu/mouse). Bladder and kidney pairs (as indicated by the labels below the x-axis) were aseptically harvested at 1 day post-infection (dpi), homogenized and plated for determination of bacterial burden (colony forming units (cfu)/ml). Mann-Whitney test was used to identify significant differences between strains in bladders and kidneys; * $p < 0.05$, n = 5 mice/strain. Each data point represents a single mouse and horizontal lines represent the median values.

S2 Fig. Deletion of UTI89 Type I RMS does not affect stationary phase gene expression. RNA sequencing comparing stationary phase transcriptomes of UTI89 and

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

UTI89*ΔhsdSMR.* Left, a volcano plot of log FDR against log fold change. Right, qq-plots showing the distribution of uncorrected p-values. Significantly differentially expressed genes (log_2 fold change >1.5 and log_{10} FDR <0.05) are labelled and colored either red (deleted genes/polar effects) or blue (validated as false positive by qRT-PCR). n = 3 biological replicates.

S3 Fig. *E. coli* strains MG1655 and CFT073 encode functional archetypal Type I RMSs with distinct specificities. (A) Transformation efficiency assay using plasmids bearing 0, 1, or 2 copies (as indicated on the x-axis) of the MG1655 Type I RMS motif (5'-AAC(N_6)GTGC-3'). Recipient cells were wild type MG1655 and the isogenic *ΔhsdSMR* mutant, as indicated by the labels below the x-axis. **(B)** Transformation efficiency assay using plasmids bearing 0, 1, or 2 copies (as indicated on the x-axis) of the CFT073 Type I RMS motif (5'-GAG(N_7)GTCA-3'). Recipient cells were wild type CFT073 and the isogenic *ΔhsdSMR* mutant, as indicated by the labels below the x-axis. Unmethylated and methylated plasmid preparations were used to transform each strain, as indicated by the legend at the top right. An unpaired t-test was used to identify significant differences between plasmids with 0, 1 and 2 methylation sites for both preparations and strains; * p <0.05, n = 3 biological replicates. Data represents mean ± s.d. of log transformed values.

S4 Fig. Loss of Type I methylation in MG1655 and CFT073 has no effect on stationary phase gene expression or growth phenotypes. (A and B) RNA sequencing comparing stationary phase transcriptomes of **(A)** MG1655 and MG1655*ΔhsdSMR*, and **(B)** CFT073 and CFT073*ΔhsdSMR.* Left, a volcano plot of log FDR against log fold change. Right, qq-plots showing the distribution of uncorrected p-values. Significantly differentially expressed genes (log_2 fold change >1.5 and log_{10} FDR <0.05) are labelled and colored either red (deleted genes/polar effects) or brown (differentially expressed genes). n = 3 biological replicates. **(C and D)** Phenotype microarray (PM) panel with plates PM1 to 20 comparing **(C)** wt MG1655 and MG1655*ΔhsdSMR*, and **(D)** CFT073 and CFT073*ΔhsdSMR.* Each plate is represented as a 12x8 grid of growth curves (red (wt), green (*ΔhsdSMR*), and yellow (overlap) on a gray background). Each growth curve plots growth (measured colorimetrically)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(y-axis) against time (x-axis). Wells representing conditions where a height difference was observed between the strains in both replicates are boxed in black, and wells which also have a quality score >150 are considered significant and boxed in red. n = 2 biological replicates.

S5 Fig. UTI89 methylation-switch strain transcriptomes are unaffected by changing Type I methylation. RNA sequencing comparing transcriptomes of corresponding growth phase wt UTI89 and **(A)** stationary phase UTI89 *hsdS*^{MG1655}, **(B)** stationary phase UTI89 *hsdSMR*^{CFT073}, **(C)** stationary phase UTI89 *hsdS*^{UTI89} , and **(D)** log phase UTI89 *hsdS*^{UTI89} Left, a volcano plot of log FDR against log fold change. Right, qq-plots showing the distribution of uncorrected p-values. Significantly differentially expressed genes ($\log_2$ fold change >1.5 and $\log_{10}$ FDR <0.05) are labelled and colored either red (deleted genes/polar effects) or brown (differentially expressed genes). n = 3 biological replicates.

S6 Fig. UTI89 Type I methylation-switch does not alter growth phenotypes. Phenotype microarray (PM) panel with plates PM1 to 20 comparing **(A)** wt UTI89 and UTI89 *hsdS*^{MG1655}, and **(B)** UTI89 and UTI89 *hsdS*^{UTI89}. Each plate is represented as a 12x8 grid of growth curves (red (wt), green (*ΔhsdSMR*), and yellow (overlap) on a gray background). Each growth curve plots growth (measured colorimetrically) (y-axis) against time (x-axis). Wells representing conditions where a height difference was observed between the strains in both replicates are boxed in black, and wells which also have a quality score >150 are considered significant and boxed in red. n = 2 biological replicates.

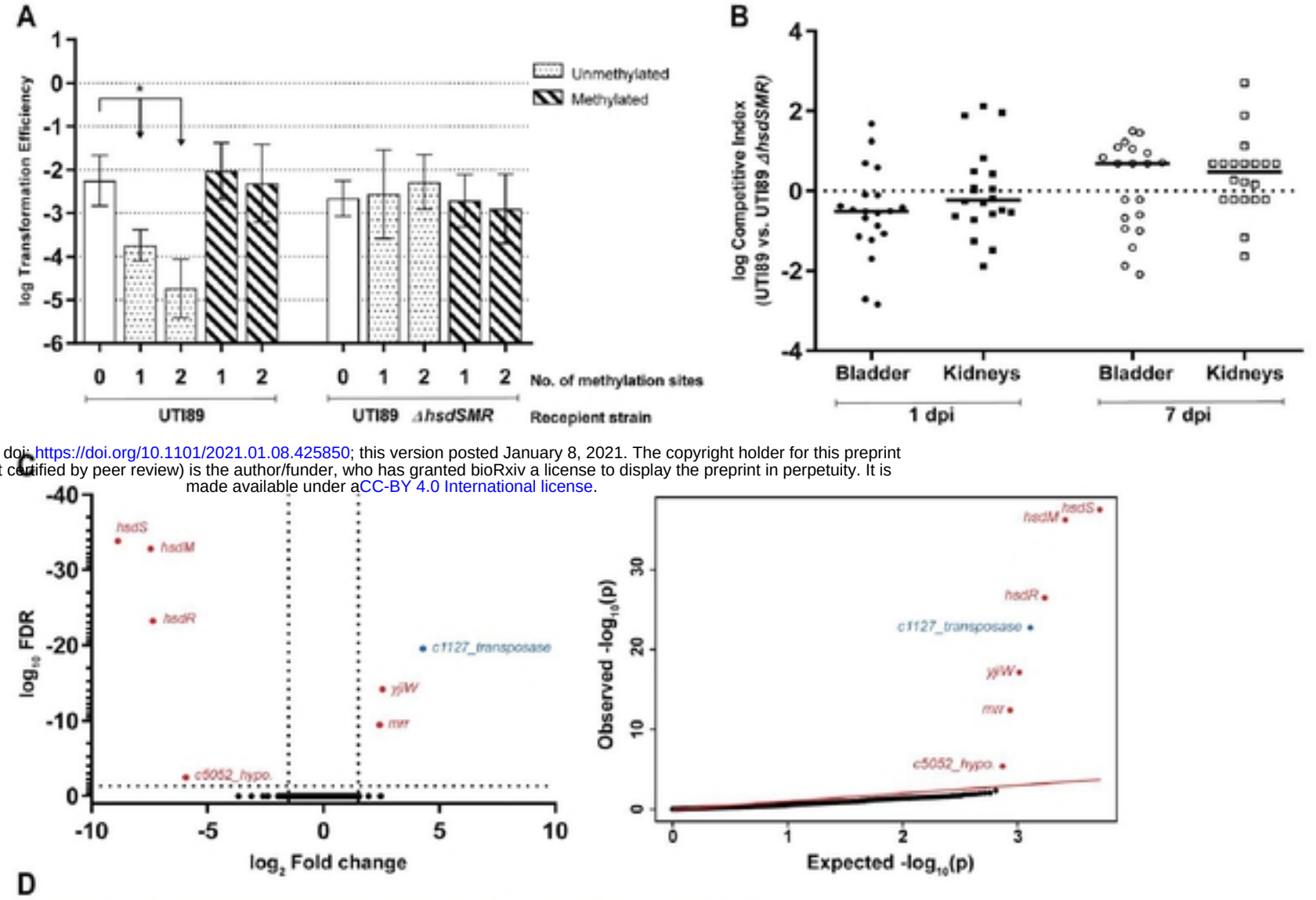
S7 Fig. *E. coli* growth is unaltered by Type I methylation. Growth curves generated using optical density (OD_{600}) values in **(A)** LB and **(B)** M9 minimal media. Wild type *E. coli* UTI89 (black), MG1655 (red), CFT073 (green), and corresponding Type I methylation mutant strains were tested, as indicated by the legend at the top right. Measurements were taken every 15 minutes and cultures propagated at 37℃. n = 3 biological replicates. Data points represent median values.

S8 Fig. Type I methylation does not affect UTI89 motility. Soft agar assay used to measure motility of UTI89 methylation mutant strains, as indicated by the legend at the top

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

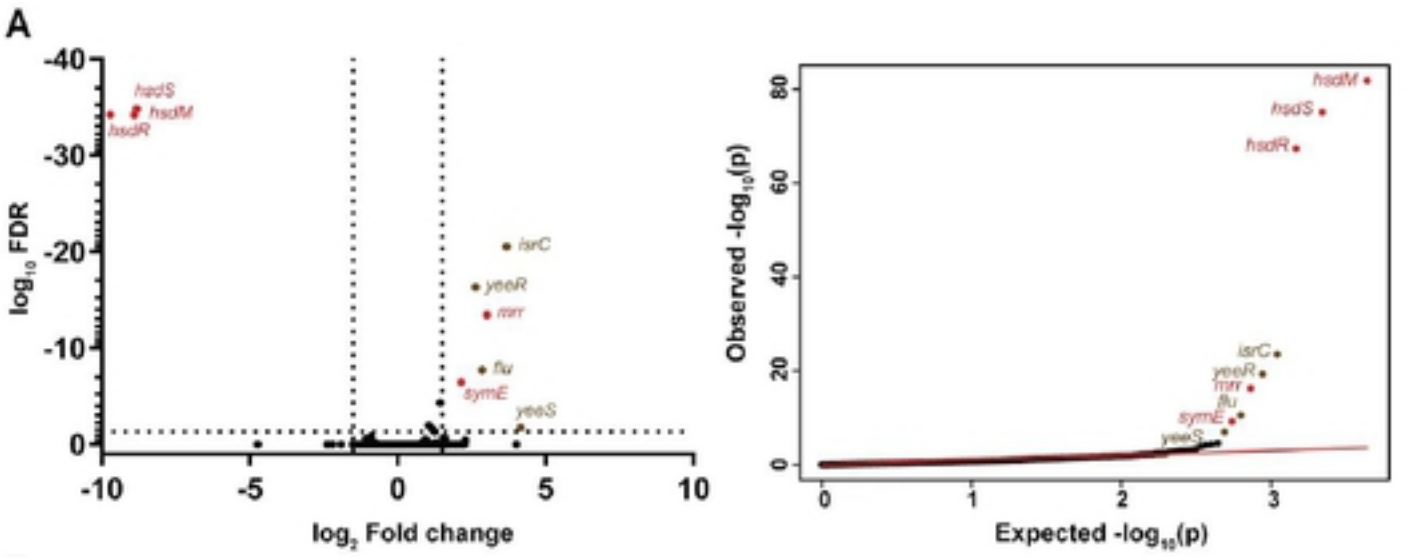
1076 right. Motility is represented relative to wild type UTI89 motility, which was included in each
1077 replicate. Strain UTI89*ΔfliC* lacking flagellin serves as a non-motile control. Wilcoxon signed
1078 rank test was used to identify significant differences in bacterial motility relative to wild type; *
1079 p <0.05, n = 3 biological replicates. Data represents median and 95% confidence intervals.
1080 **S9 Fig. UTI89 biofilm formation is not affected by Type I methylation.** 96-well crystal
1081 violet assay for biofilm quantification using LB **(A and B)** or YESCA **(C and D)** media. UTI89
1082 Type I methylation mutant strains were tested, as indicated by the legend at the top right.
1083 Assay was performed for 24 hours **(A and C)** or 48 hours **(B and D)**, as well as at 26℃ and
1084 37℃ for each media as indicated by the labels below the x-axis. Biofilms were quantified
1085 using crystal violet absorbance at 590nm (OD_{590}) and represented relative to wild type UTI89
1086 which was included in each replicate and condition. Wilcoxon signed rank test was used to
1087 identify significant differences in biofilm formation relative to wild type; * p <0.05, n = 3
1088 biological replicates. Data represents median and 95% confidence intervals.
1089 **S10 Fig. MG1655 Type I methylation-switch has no effect on growth phenotypes.**
1090 Phenotype microarray (PM) panel with plates PM1 to 20 comparing **(A)** wt MG1655 and
1091 MG1655 $hsdS^{UTI89}$, and **(B)** MG1655 and MG1655 $hsdS^{MG1655}$. Each plate is represented as
1092 a 12x8 grid of growth curves (red (wt), green (*ΔhsdSMR*), and yellow (overlap) on a gray
1093 background). Each growth curve plots growth (measured colorimetrically) (y-axis) against
1094 time (x-axis). Wells representing conditions where a height difference was observed
1095 between the strains in both replicates are boxed in black, and wells which also have a quality
1096 score >150 are considered significant and boxed in red. n = 2 biological replicates.

Fig 1.



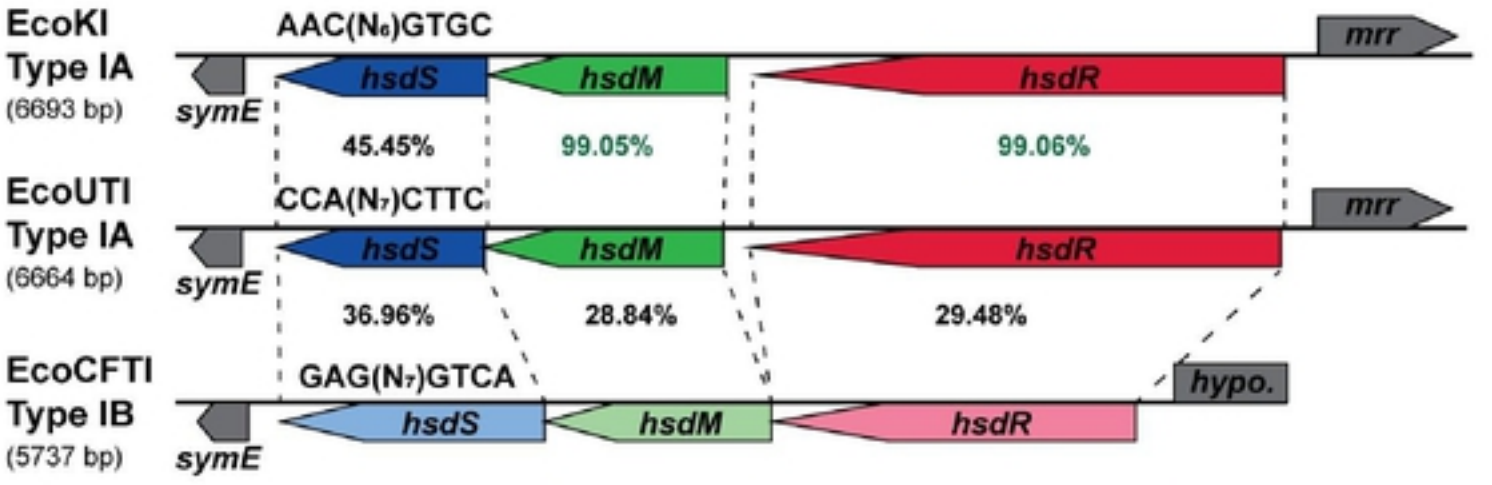
bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig 2.



bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Fig 3.



bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

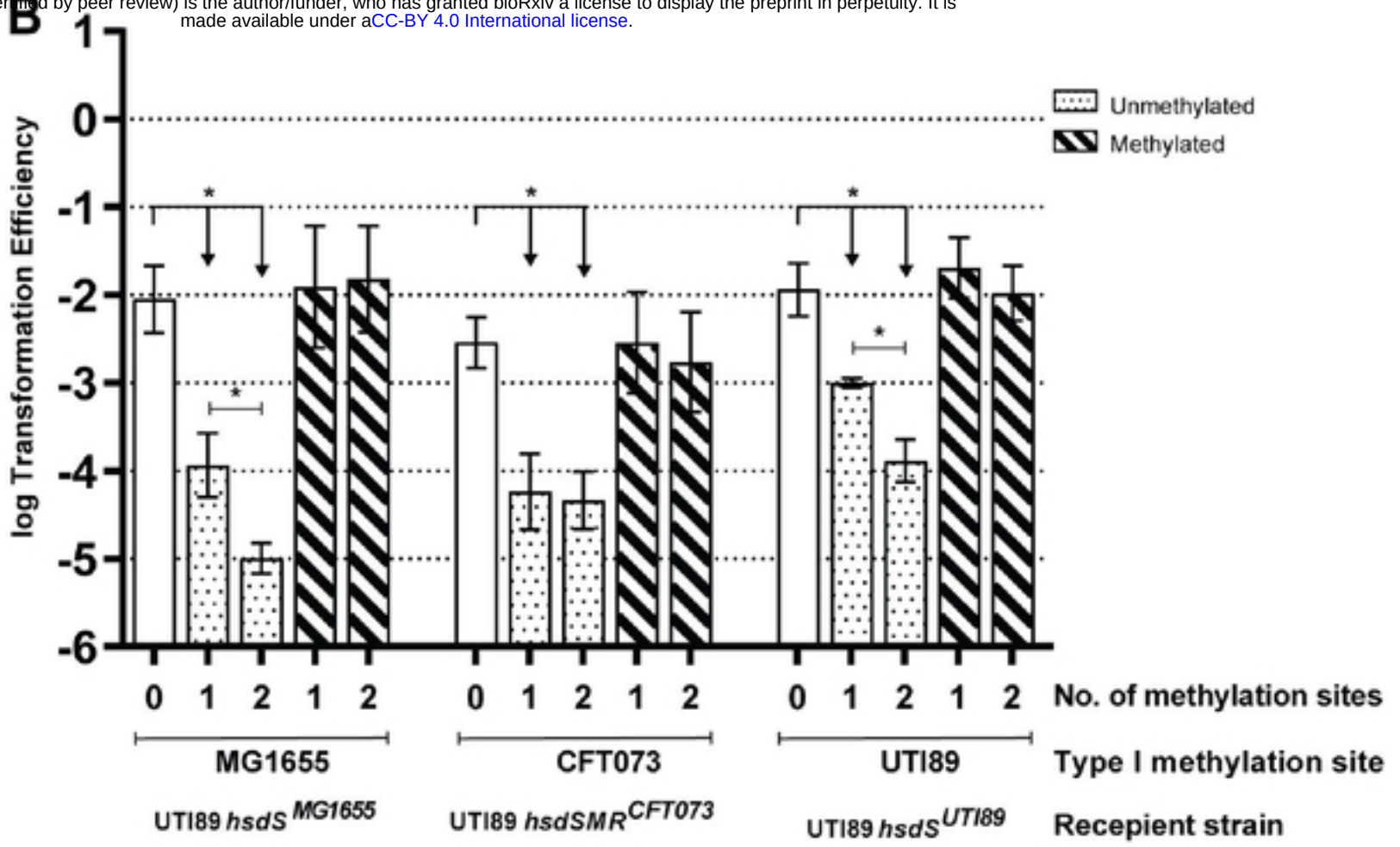
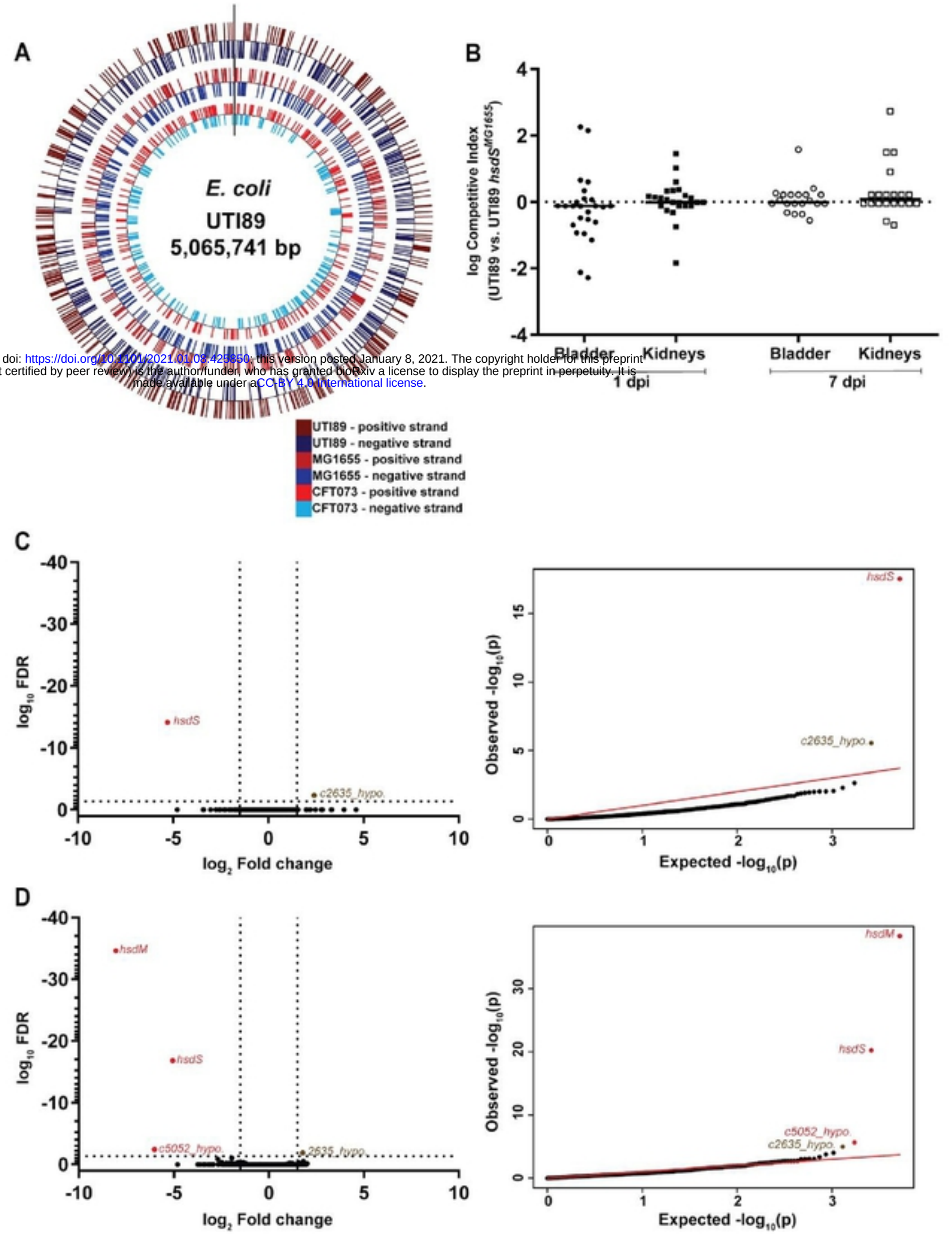


Fig 4.



bioRxiv preprint doi: https://doi.org/10.1101/2021.01.08.425850; this version posted January 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.