1   **Reassembling a cannon in the DNA defense arsenal: genetics of StySA, a BREX**

2   **phage exclusion system in *Salmonella* lab strains**

3

4   Julie Zaworski*, Oyut Dagva*, Julius Brandt*, Chloé Baum*, Laurence Ettwiller*,

5   Alexey Fomenkov* and Elisabeth A. Raleigh*

6   * Research Department, New England Biolabs, Ipswich, MA 01938-2723

7

8   Running Head: *Salmonella* BREX phage exclusion system

12

13   Corresponding author:

14   Elisabeth A. Raleigh

15   Research Department

16   New England Biolabs

17   240 County Road

18   Ipswich, MA 01935

19   978 380 7238

20   raleigh@neb.com

Abstract

Understanding mechanisms that shape horizontal exchange in prokaryotes is a key problem in biology. A major limit on DNA entry is imposed by restriction-modification (RM) processes that depend on the pattern of DNA modification at host-specified sites. In classical RM, endonucleolytic DNA cleavage follows detection of unprotected sites on entering DNA. Recent investigation has uncovered BREX systems, RM-like activities that employ host protection by DNA modification but replication arrest without evident nuclease action on unmodified phage DNA. We show that the historical *stySA* RM locus of *Salmonella enterica* sv Typhimurium is a BREX homolog. The *stySA29* allele of the hybrid strain LB5000 carries a mutated version of the ancestral LT2 BREX system. Surprisingly, both a restriction and a methylation defect are observed for this lineage despite lack of mutations in *brxX,* the modification gene homolog. Instead, flanking genes *pglZ* and *brxC* each carry multiple mutations (μ) in C-terminal domains. To avoid plasmid artifacts and potential stoichiometric interference, we chose to investigate this system *in situ*, replacing the mutated *pglZμ* and *brxCμ* genes with wild type (WT). PglZ-WT supports methylation in the presence of either BrxCμ or BrxC-WT but not in the presence of a deletion/insertion allele, Δ*brxC::cat.* Restriction of phage L requires both BrxC-WT and PglZ-WT, implicating the BrxC C-terminus specifically in restriction activity. Disruption of four other CDS with *cat* cassettes still permitted modification, suggesting that BrxC, PglZ and BrxX are principal components of the modification activity. BrxL is required for restriction only. A partial disruption of *brxL* disrupts transcription globally.

Introduction

44   Transfer of genes from one organism to another shapes ecological capacities in

45   microbiomes on both short and long-term timescales. Thus, mechanisms that limit or

46   promote such transfer are of fundamental interest. Ecologic interactions with phage play

47   a major role in host colonization by prokaryotes [1, 2]. Prokaryote defenses against

48   phages are of particular interest [3, 4], particularly as therapeutic uses are contemplated

49   [5]. Microbiome and metagenome studies have led to a renaissance in the study of

50   phage-host interaction [6, 7].

51   Host defensive processes include restriction-modification (RM) systems as major

52   contributors[8]. These distinguish the host DNA from foreign invaders using the pattern

53   of DNA modification; DNA with the wrong modification pattern is rejected [9, 10]. DNA

54   modification by bacterial hosts is typically carried out by methylation of adenine or

55   cytosine bases. Recently, sequence-specific protective phosphorothioation of the DNA

56   backbone has emerged as an important biological phenomenon [11]. In most familiar

57   cases of RM systems, protection is conferred by methylation of a particular base within a

58   specific sequence motif, while rejection consists of endonucleolytic cleavage of both

59   strands in response to the presence of unmethylated specific sites. Such cleavage leads to

60   very rapid interruption of the phage development program.

61   Both defensive and epigenetic processes can involve DNA modification states, so

62   taxa with no DNA modification are extremely rare. Epigenetic regulation is important in

63   the life of the cell, and often the relevant genes are fixed in a lineage, while defense

64   functions are diverse [12, 13].

65   A further result of phage-bacterial interaction are countermeasures: phage-or

66   plasmid-borne antirestriction functions [14-19]. In this regard, we note that *in vivo*, Iida

67   and coworkers found the StySA restriction activity to be sensitive to the antirestriction

68   activity of EcoP1 DarA [20], an activity otherwise only know to affect Type I restriction.

69   Extension of Dar susceptibility to non-Type I enzymes would shed light on the scope of

70   protection afforded by this phage countermeasure.

71   These defense mechanisms often cluster in genome islands. Genome comparisons

72   between close prokaryote relatives led to the definition of variable "genome islands",

73   which are clusters of genes in one genome with no counterpart in another [6, 8, 21].

74   These variable genes usually deal with changing environments e.g. for resource

75   acquisition and biological interactions. Some genome islands are specialized "defense

76   islands", enriched in genes that specifically regulate DNA entry [8, 22-26]. Defenses

77   include RM systems (including modification-dependent nucleases, MDRS), and a wide

78   variety of "abortive infection" elements. Abortive infection (Abi) processes prevent

79   phage transmission to sister cells via cell death [27]. Abi mechanisms mostly target later

80   events in phage development; even virion expression can occur (e.g. [28]). Defense

81   islands may carry elements related to the Pgl (phage growth limitation) system in

82   *Streptomyces coelicolor*, studied by M.C. Smith and coworkers [29-31], in which PglX-

83   mediated methylation conditions the phage to restriction by MDRS upon infection of

84   sibling cells. In this system the host genome is *not* methylated.

85   Sorek and co-workers extended the suite of methylation-protected defense

86   systems using neighborhood analysis anchored by homologs of a component of the Pgl

87   system, PglZ [32, 33]. This identified a set of systems designated BREX (Bacteriophage

88   Exclusion), gene clusters of 4 to 8 genes, depending on the subtype. A BREX system

89   from *Bacillus cereus* was studied experimentally in *B. subtilis* by Goldfarb et al.[32].

90   Further experimental characterizations include a system in *E. coli* strain HS [16, 34]; one

91   in *Lactobacillus casei* [35]; and a plasmid-borne system from *Escherichia fergusoni* [36].

92   Though some elements of BREX are related to Pgl, the two families displayed important

93   differences in biological endpoints, particularly the role of methylation, which protects

94   BREX hosts [34] but elicits restriction by Pgl sibling cells.

95   SenLT2II (StySA) is one of three RM systems in *S. enterica* sv Typhimurium

96   LT2. The other two are multicomponent ATP-dependent systems of Type I (SenLT2I;

97   LT, StyLT in the early literature) and Type III (SenLT2III; SB, StySB). SenLT2II

98   (StySA) was shown to carry out modification of adenine residues [37]. Its distribution

99   may be limited to serovar Typhimurium; a survey of 85 other *Salmonella* serovars that

100   included 11 testable for StySA found no SenLT2II (StySA) restriction activity outside of

101   serovar Typhimurium [38]. For clarity we use the terminology from the restriction

102   literature here: StySA, which is SenLT2II in the reference LT2 sequence.

103   We identified the genomic location of the StySA system while analyzing the

104   sequences of *S.* Typhimurium-*S.* Abony hybrid strains[39]. The triply restriction-deficient

105   laboratory shuttle host LB5000 was constructed to be a restriction-deficient but

106   modification-proficient recipient for molecular genetic manipulations [40]. LB5000 (our

107   STK005) carries a mutated allele of StySA (SenLT2II) named *stySA29*. This allele is also

108   carried by our experimental host ER3625, descended from LB5000 [39]. LT2 locus_ID

109   STM4495 is located genetically in the appropriate place to specify StySA. The single

110   large protein specified by STM4495 was initially predicted to carry both R and M

111   activities (REBASE 2015; 1225 aa; Type IIGS gamma). This is quite distinct from Type I

112   enzymes, which require two genes for M activity and a third for R activity. Notably, this

113   locus_ID is within a variable chromosomal island anchored by *leuX* where numerous

114   non-homologous mobile elements are found in *E. coli* and *Salmonella* [41].

115   In this study, we describe the relationship of the StySA RM system to BREX

116   systems and report the genetic contributions of its constituent genes to host protection

117   and interruption of phage infection. The transcriptional organization of the locus

118   contributes to understanding of relative expression levels in the native context. We find

119   that *brxC* and *pglZ* mutations in our LB5000 descendant are responsible for the lack of

120   restriction and modification; only the N-terminal domain of BrxC is required for

121   modification. Serendipitously, we present evidence that the BrxL C-terminal domain by

122   itself can have a large effect on host gene expression, particularly a non-SOS effect on

123   resident prophages.

124   **Results**

125   **Strain engineering design: domains, mutation clusters and transcription start sites**

126   StySA is a BREX homolog

127   The non-restricting strain LB5000 and our experimental host ER3625 carried

128   LT2-like m6A modification patterns, with one surprising exception: no modification

129   attributable to STM4495 was found [39] despite identity of locus_ID JJB80_22595 to the

130   LT2 sequence. Flanking genes corresponding to STM4492 and STM4496 did vary from

131   the LT2 sequence (see further below; for a guide to terminology, see Materials and

132   Methods).

133   Automated annotation [42] predicts a relationship of the StySA region with

134   BREX and Pgl systems; we have adopted the automated name assignments from NCBI.

135   The BREX homology is compelling when aligning the StySA region of LT2 (from [43]))

136   with that of the *E. coli* HS BREX locus [34] (Figure 1A). In addition to the 6 genes

137   described for the *E. coli* and *B. subtilis* BREX type I gene clusters, the LT2 StySA/BREX

138   region carries an extra pair of genes, STM4493 and STM4494 (STM4493-4). STM4493-

139   4 interrupt the similarity of the gene cluster (in the alignment gap Figure 1A). STM4493-

140   4, which we will refer to as "accessory genes", were apparently acquired by *Salmonella*

141   separately from the genes we will designate "BREX homologs" (see Fig. 1 of [41]). We

142   find below that neither of the accessory genes contributes to modification activity *in vivo*

143   (see below "Phenotypic consequences of strain engineering").
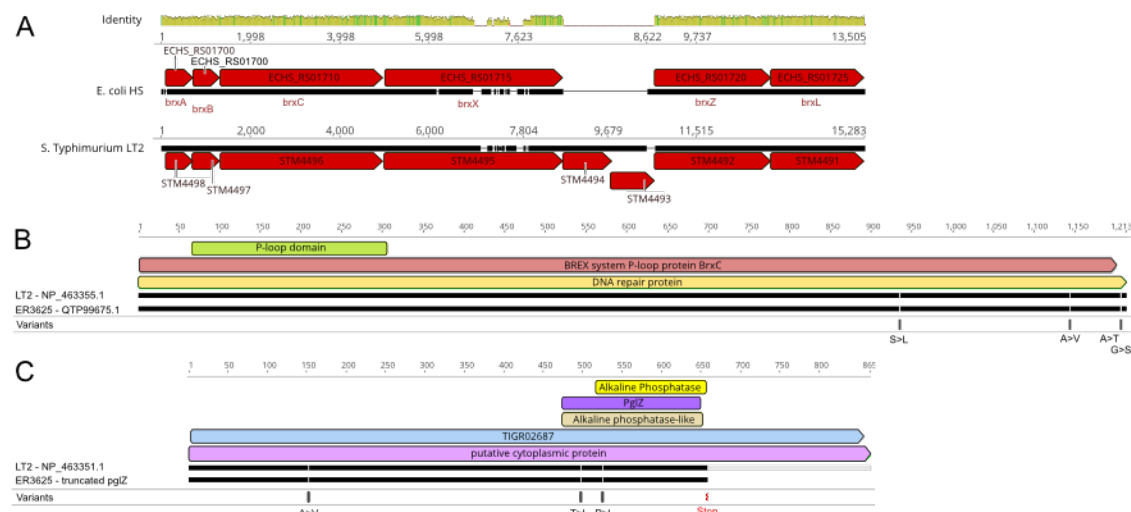
144

145   Figure 1: StySA locus similarity to *E. coli* BREX and protein variants in ER3625. Panel A, DNA alignment of

146   BREX regions from LT2 with *E. coli* HS. Identity: green 100%, yellow close similarity, white sequence

147   interruption. Numbers are nucleotide coordinates of segments extracted from the genome sequences (HS,

148   LT2). Red boxes are annotated CDS with locus_ID numbers; black segments below the arrows represent

149   DNA, with breaks where the two don't match. A fine point is the small gap in the LT2 black line just

150   upstream of *brxZ*: STM4491 and ECHS_RS01720. The black line for E. coli has a bit of sequence that has no

151   counterpart in LT2. Presumably in E. coli there is a promoter/regulatory region for *brxZ*. For LT2 this

152   function may be provided by 3' end of STM4493 DNA. Panels B (BrxC) and C (PglZ) display amino acid

153   coordinates for proteins of LT2 and its descendant ER3625. Breaks in the black lines between signify

154   variant amino acid positions; the grey extension for the LT2 PglZ represents the C-terminal region missing

155   in ER3625 (i.e. after the point of the translation termination for *pglZμ*). Colored boxes show predicted

156   domains annotated (see text). Variant amino acids resulting from mutations present in ER3625 as

157   compared to LT2 are indicated in the bottom rows of B and C.

158

159   Mutation set leading to StySA R$^-$M$^-$ phenotype for *stySA29*

160   ER3625 genes descended from STM4492 and STM4496 will be designated

161   *brxCμ* (JJB80_22600, coding for a mutated version of BrxC) and *pglZμ* (JJB80_22580,

162   coding for a mutated version of PglZ) respectively. Each carries multiple mutations

163   relative to the LT2 ancestor. The amino-acid changes resulting from these are displayed

164   in Figure 1B (BrxC) and Figure 1C (PglZ). The multiplicity of changes we attribute to the

165   use of nitrosoguanidine mutagenesis during isolation of the *stySA29* allele [39, 44]. The

166   R$^-$M$^+$ phenotype originally reported for the lineage [40] was unstable, losing modification

167   (M$^+$) ability [45]. Since tandem duplications are frequent in *Salmonella* [46], the original

168   strain might have carried a duplication of the region, with clustered mutations in different

169   genes in each copy, followed by resolution to single copy with both genes mutated and an

170   R$^-$M$^-$ phenotype. Our genetic analysis below (Phenotypic consequences of strain

171   engineering) is compatible with this scenario.

172   Genetic approach to the StySA locus

173   We chose to investigate this system *in situ*, replacing the mutated *pglZ* and *brxC*

174   genes with wild type (WT). The engineering strategy first required a sketch of possible

175   transcription signals, due to the complex organization of the locus and multiple potential

176   toxic interactions. The strategy for gene deletion and replacement employed a method

177 that leaves no scars, via an intermediate carrying a drug resistance cassette with its own

178 promoter (see Materials and Methods and S2 Fig.). Strains created are listed in S2 File.

179 Transcription overview: StySA operon structure with Cappable-Seq and RNAseq

180 We first investigated transcription start sites with Cappable-seq [47] in the

181 ancestral LT2 strain. The goal of such experiments is to better comprehend the complex

182 organization of the BREX locus and the potential toxic interactions. It will also facilitate

183 the design of deletion strains (see below). In our lineage with the reconstructed wild type

184 StySA locus (JZ_058), we assessed the landscape of transcription using RNA-seq.

185 With Cappable-seq, two biological replicates identified 15,650 and 15,145 unique

186 TSS positions in the genome (Material and Methods) (above the TPM >= 1.0 and

187 EnrichRatio >= 2.5). Clustering was used to regroup close TSS giving a final count of

188 9,422 and 9,777 TSS clusters. Of these, 8,041 are shared positions with an adjusted R2 =

189 0.96426 and a P-value < 2.2E^{-16}.

190 With this highly accurate method for mapping TSS location genome-wide, we

191 confirm in our LT2 isolate (STK013) some TSS already made publicly available [48] and

192 identify some new intragenic TSS (Figure 2). Primary TSS upstream of STM4498 (*brxA*)

193 and STM4492 (*brxZ*) are confirmed; internal TSS are found within the STM4495 (*pglX*)

194 and STM4494 (DUF4495) CDSs. A strong TSS cluster is located upstream of STM4490

195 (*mrr2*) in the forward orientation; this is adjacent to, but outside of the BREX/StySA

196 cluster, defined as STM4491 (*brxL*)-STM4498 (*brxA*).

197 RNASeq experiments allow us to map transcript profile and abundance (TPM)

198 across the StySA locus in JZ_058, which has restored the locus to sequence identity with

- 9 -

199   LT2 in the ER3625 lineage. We observe agreement between transcript abundance for

200   each feature and available TSS for the ancestral LT2 region in STK013: the RNA profiles

201   along the locus show varying transcript levels with alternating regions of higher and

202   lower abundance. We infer likely transcription starts directly upstream of STM4498

203   (*brxA*), and within STM4494 (DUF4495) and STM4492 (*pglZ*) homologs. For the TSS

204   experiment, we observed the TSS in front of STM4496 (*brxC*) was weaker than that in

205   front of STM4498 (*brxA*); the TPM data suggest that this *brxC* promoter could be as

206   strong as the others in this strain background.
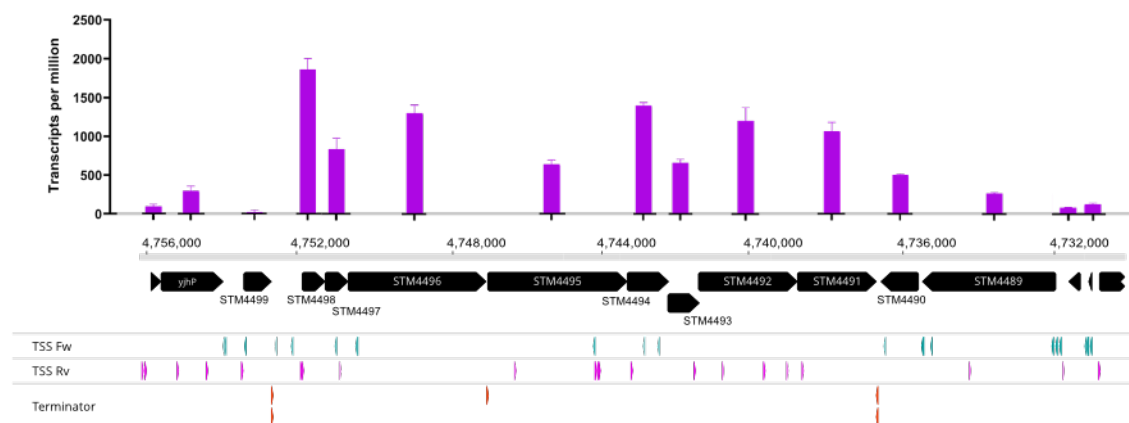
207

208   Figure 2: StySA locus operon structure analysis. Top panel corresponds to transcripts per million for the

209   WT StySA (JZ_058) strain. Abcissa coordinates refer to LT2 sequence NC_003197.2. Black arrows are CDS

210   (to scale), labelled with the locus_ID of LT2. The BREX-related genes are STM4498-4491; on the flanks,

211   three (left) and five (right) external CDS are included to enable estimation of transcription reading into or

212   out of the locus. The bottom panel show transcription start sites (TSS) determined in LT2 wild type using

213   Cappable-seq. "Fw" is forward with respect to the genome coordinates, not the reading frames; "Rv"

214   reverse. Transcription terminators were predicted using TransTerm.

215   Rho-independent terminators are predicted at three positions in this locus. Two

216   are bidirectional, flanking the StySA locus (before STM4498 (*brxA*) and after STM4491

217   (*brxL*)). The locus should thus be transcriptionally insulated (see **"Local and global**

218   **inferences from transcriptome profiling" belo*w*).**

219   The third predicted terminator lies between STM4496 (*brxC*) and STM4495

220   (*pglX*), suggesting multiple operons within the locus. Experimental evidence for the *brxC*

221   terminator was sought, using a combination of SMRT-Cappable-seq and RACE methods

222   (S9 File). The combination approach (S9 panel A) supports the authenticity of the *brxC*

223   transcription terminator: read number drops at the end of *brxC* (S9 panels B and C)*,* with

224   some read-through to *brxX*. Coverage was too low for stronger conclusions (~15 reads

225   before the terminator position and ~5 after it).

226   These experiments suggest that STM4494, STM4493, *pglZ* and most likely *brxL*

227   are on one transcript (S9 panel D), while *brxA*, *B* and *C* and *pglX* are on a different

228   transcript (S9 panel A).

229   <u>Putative functions of the BREX core gene components and response to genome</u>

230   <u>manipulation</u>

231   For each annotated CDS in the ER3625 StySA locus, gene segments were deleted

232   with consideration of the transcript structures predicted from the analysis above. The

233   engineering strategy for gene deletion and replacement employed a method that leaves no

234   scars, via an intermediate carrying a drug resistance cassette with its own promoter (see

235   Materials and Methods and S2 Fig.). Strains created are listed in S2 File

236   Functional domain annotations were described for the Pgl and BREX systems.

237   The characterized Pgl system expresses a putative phosphatase (*pglZ*), a protein kinase

238   (*pglW*), a candidate adenine-specific DNA methyltransferase (*pglX*) and a P-loop ATPase

- 11 -

239   (*pglY*) [29]. Proteins specified by two of the 6 genes in type I BREX are similar: PglZ

240   and PglX. BREX-specific genes specify BrxL, identified as a Lon-like protease-domain;

241   BrxA, proposed to be a structural homologue of RNA-binding antitermination protein

242   NusB; BrxB, a protein of unknown function and BrxC, a large protein with a P-loop

243   ATP-binding domain [32]..

244   Schematic domain predictions are shown in Figure 1B (*brxC*, STM4496 in LT2;

245   JJB80_22600 in ER3625), Figure 1C (*pglZ*, STM4492 in LT2; JJB80_22580 in ER3625)

246   and in S1 File (*brxA*; STM4498 and JJB80_22610; *brxB*, STM4497 and JJB80_22605;

247   *pglX*, STM4495 and JJB80_22595; STM4494 (JJB80_22590), STM4493 (JJB80_22585)

248   and *brxL*, STM4491 and JJB80_22575).

249   From left to right in Figure 1A, STM4498 codes for a BrxA homolog. The gene

250   contains a putative RBS for STM4497. Domain prediction of DUF1819, is inferred to

251   code for an inner membrane protein. We were unable to create a *cat* replacement of this

252   gene.

253   STM4497 codes for a BrxB homolog, annotated as a DUF1788 protein.

254   Replacement of *brxB* with *cat* has extensive effects on transcription of the cluster. The

255   *cat* replacement is viable in three allelic states (see "Local and global inferences from

256   transcriptome profiling" below).

257   STM4496 codes for a BrxC/PglY homolog, originally annotated in LT2 as

258   "putative ATPase involved in DNA repair". Geneious implementation of InterProScan

259   with LT2 protein sequence AAL23314 (NP_463355.1) yielded domain hits including an

260   N-terminal ATPase region (P-loop NTPase, ATPase involved in DNA repair, P-loop

261   containing nucleoside triphosphate hydrolase) and a C terminal region designated "SMC

262   N-terminal domain". In ER3625, the four mutations in *brxCµ* (JJB80_22600) are

263   clustered in the 3' end of the gene (Figure 1B). InterProScan detects only the NTPase hits

264   in the N-terminus in the corresponding protein (QTP99675.1). Presumably the variant

265   amino acids result in loss of the SMC-domain recognition.

266   STM4495 is a BrxX/PglX homologous gene, with signatures of m6A

267   methyltransferases. Domain and annotation search leads to N-6 adenine-specific DNA

268   methylases hits. DNA alignment with the *E. coli* HS homolog show two patches of

269   divergence (Figure 1A). This is compatible with potential DNA recognition in that

270   region, since the two MTase recognize and modify different motifs. Methyltransferases

271   often segregate site specificity into a recognition module [49-53]. We were unable to

272   make a deletion of this gene in three different allelic states. Toxicity due to disruption of

273   transcription in the intermediate or a toxic protein complex formed from other

274   components in its absence are possible explanations.

275   STM4492 is a BrxZ/PglZ homolog, thus putatively has phosphatase activity. PglZ

276   of *Streptomyces coelicolor* was required for viability in the presence of PglX; neither

277   gene deletion nor point mutation of the candidate catalytic aspartates in the phosphatase

278   fold could be made in single copy [29]. Geneious implementation of InterProScan with

279   LT2 protein sequence AAL23310 (NP_463351.1) yielded hits with four different

280   alkaline-phosphatase-like domains; these are found in alkaline phosphatases,

281   arylsulphatases, phosphoglycerate mutase, phosphonoacetate hydrolases or

282   phosphoenolmutases.

283   The mutated homolog in ER3625 (*pglZµ*; JJB80_22580) carries five mutations.

284   Notable is a stop codon just 3' to the annotated PglZ domain. The N-terminal protein

- 13 -

285   fragment (here designated PglZ') would carry three variant amino acids when translated;

286   a fifth mutation is distal to the stop. No protein corresponding to PglZ' is listed in the

287   NCBI proteome for ER3625, although the gene is annotated. To determine whether PglZ'

288   is recognized by domain-detection programs, a translation of the annotated gene

289   (truncated at the stop) was submitted to InterProScan. The PglZ annotation was returned

290   to this search (Figure 1C); thus the structural domain PglZ/AlkPhos is recognized even

291   when the tail of the protein is not there. This does not prove functionality but suggests

292   that partial assembly or partial enzymatic function could be retained. The intact gene is

293   required for both R and M activity *in vivo* (see below). In addition, we have strong

294   evidence that the stop codon is polar on downstream transcription.

295   STM4491 (JJB80_22575) is a BrxL homolog. Domain predictions for protein

296   sequence AAL23309 (NP_463350.1) gave multiple hits to Lon-protease domains. This is

297   proposed to be two-domain Lon-like protease structure with an N-terminal ATPase and a

298   C-terminal proteolytic domain. However, it is most likely not a protease as the catalytic

299   site is not present (see further below).

300   **Phenotypic consequences of genome engineering.**

301   Five phenotypic measurements were carried out with the strains constructed:

302   phage restriction; modification at StySA sites (using Dam modification as a control);

303   growth rate; expression level of the M gene (*pglX,* STM4495; JJB80_22595); and for a

304   selection of strains, RNAseq was used to analyze local and global transcription effects.

305   Colony formation was found to be slow and variable on plates for some strains

306   and not others (See S3 Fig.). Thus, a comparison of growth rate in liquid was undertaken.

307   PglZ and BrxC variants contribute to both R$^-$ and M$^-$ properties

308   Only one R$^+$ strain was obtained. JZ_058 was able to restrict phage L; it carries

309   wild-type LT2 sequence in place of both *brxC* and *pglZ* of ER3625. The magnitude of

310   restriction agreed with literature reports [40, 54]: 100-fold reduction in plaque-forming

311   ability (S3 Fig.).

312

A

B — Effect of BrxC alleles on methylation
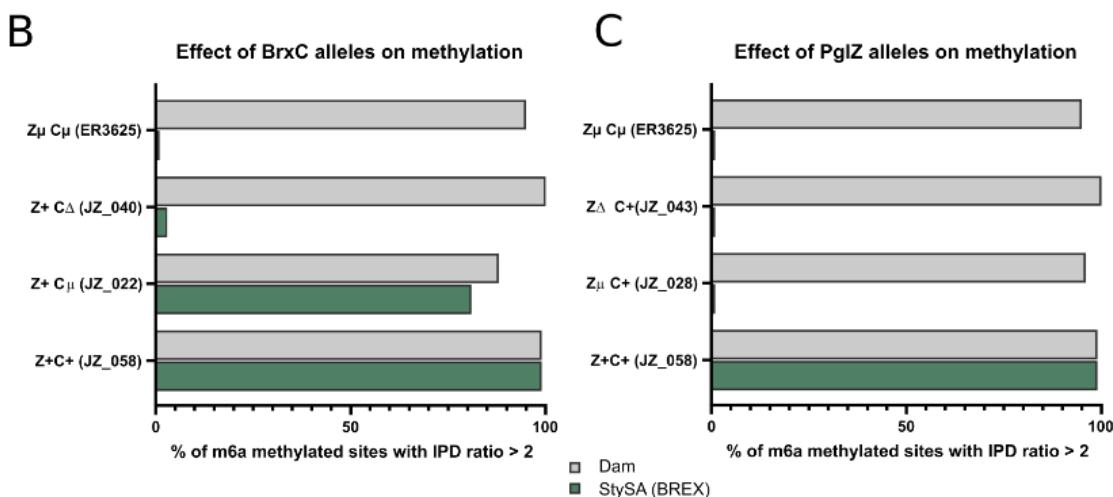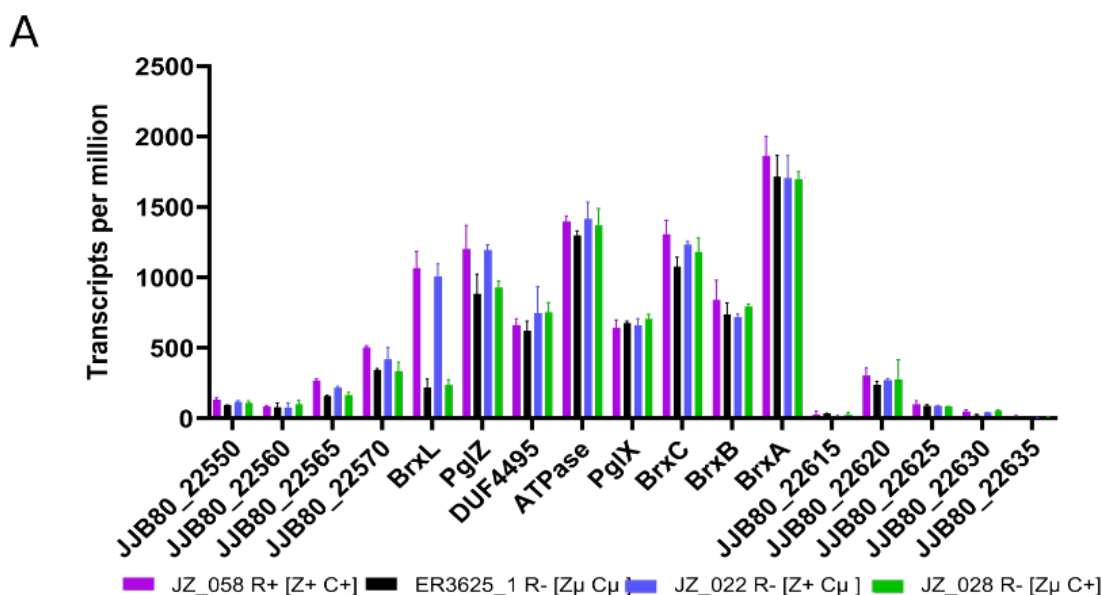
C — Effect of PglZ alleles on methylation

313 Figure 3: BrxC and PglZ alleles determine restriction and methylation phenotypes, and BrxL transcript

314 abundance. Panel A: Transcription level across features in the StySA gene neighborhood for four strains

315 with clustered mutations (μ) or wild type (+) in *brxC* and *pglZ*. Strain restriction phenotypes are R$^+$: 100-

316 fold reduction in plaque formation with phage L; R$^-$ no restriction activity. Features shown (in opposite

317 orientation to Figures 1 and 2; not to scale) are NCBI-assigned locus_ID (outside the StySA cluster) or gene

318 names (within it); vertical axis is transcripts assigned per million transcripts sequenced. *pglZμ* is polar on

319 *brxL*. Panels B and C: effects of *brxC* (Panel B) and *pglZ* (Panel C) alleles on methylation of StySA sites

320 (green bars) with Dam sites (gray bars) as a control.

321 Modification *in vivo* did not require wild-type BrxC: when *pglZ* is wild-type,

322 mutated *brxCμ* allows modification (Figure 3 panel B, JZ_022). However, the N-

323 terminus of BrxC is required for modification, since Δ*brxC::cat* no longer modifies

324 (Figure 3 panel B, JZ_040). In contrast, modification required PglZ to be intact (Figure 3

325 panel C: JZ_028 and JZ_043).

326 Methylation activity does not depend on the level of PglX (MTase) expression

327 We used qPCR to measure *pglX* transcription levels to clarify the question of

328 whether the degree of methylation is regulated by changes in transcription that result

329 from engineering steps. The *pglX* transcript level does not correlate with M phenotype

330 (**Error! Reference source not found.** panel B top, M$^+$, StySA modified; M$^-$ StySA

331 unmodified; M^p, partial modification). Significant differences in transcription are seen in

332 some strains relative to ancestor ER3625, but the highest transcription is found in

333 unmodified strains. A negative control is OD_127, a strain with a multigene deletion

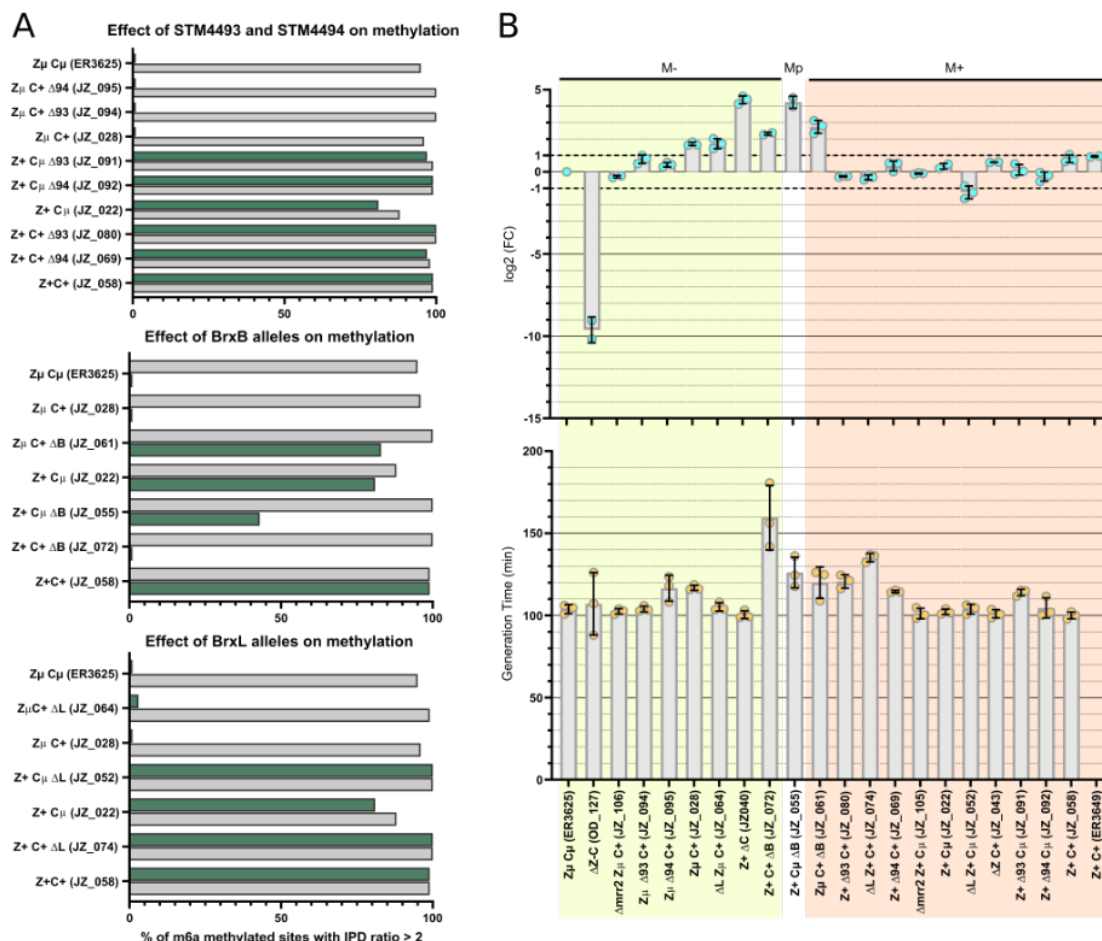334 removing *pglZ-brxC* and including *pglX*.

335

Figure 4: Modification, PglX transcript and growth rate responses to engineering. Panel A, Methylation of Dam sites (gray bar, internal control) and StySA sites (green bars) with different genetic configurations. μ, mutation cluster inherited from LB5000; +, wild type LT2 replacement; Δ, segment of the gene replaced with a *cam* cassette; (), strain number. Panel B, STM4495 (*pglX*) expression in the different strains. Each dot represents a biological replicate, the bar chart represents the mean relative expression level. The methylation level is simplified and represented by M-, Mp and M+ respectively for methylation < 5%, methylation at an intermediate level and methylation > 90%. Panel C, Generation time in minutes. Each dot represents a biological replicate, the bar chart represents the mean relative expression level.

Methylation activity is not affected by ΔSTM4493::*cat*, ΔSTM4494::*cat* or Δ*brxL::cat*

Removal of accessory genes STM4493 or STM4494 do not affect methylation

(Figure 4 panel A top). M⁺ *pglZ⁺* (*brxCμ* or *brxC⁺*) hosts retain modification with

ΔSTM4493::*cat* (JZ_091 or JZ_080) or ΔSTM4494::*ca*t (JZ_092 or JZ_069);

- 17 -

348   modification did not return to the M⁻ *pglZμ brxC⁺* host (JZ_094 and JZ_095). Similarly,

349   removal of the N-terminal segment of *brxL* did not change M status (Figure 4 panel A

350   bottom). The contradictory properties of Δ*B::cat* are discussed further below.

351   Growth rate is affected by Δ::*cat* constructions

352   We found on plates that the strains have different colony phenotypes and different

353   growth rates (See S3 Fig.) so we did a liquid culture experiment in plates to accurately

354   compare the strains growth, see Figure 4 panel B bottom. Most strains show growth rates

355   about the same as the ancestor ER3625. Interestingly, there is no correlation between

356   methylation level and loss of fitness. However, the strains with Δ*brxB::cat* (JZ_072,

357   JZ_055 and JZ_061) share slower growth rate, regardless of the allelic state of *brxC* and

358   *pglZ*. We attributed this shared property to the unbalanced expression of the whole

359   operon due to the *cat* cassette promotor, see below and Figure 5. Of the three Δ*brxL::cat*

360   strains, the one closest to wildtype also grows more slowly: JZ_074 (Δ*brxL::cat pglZ⁺*

361   *brxC⁺*).

362
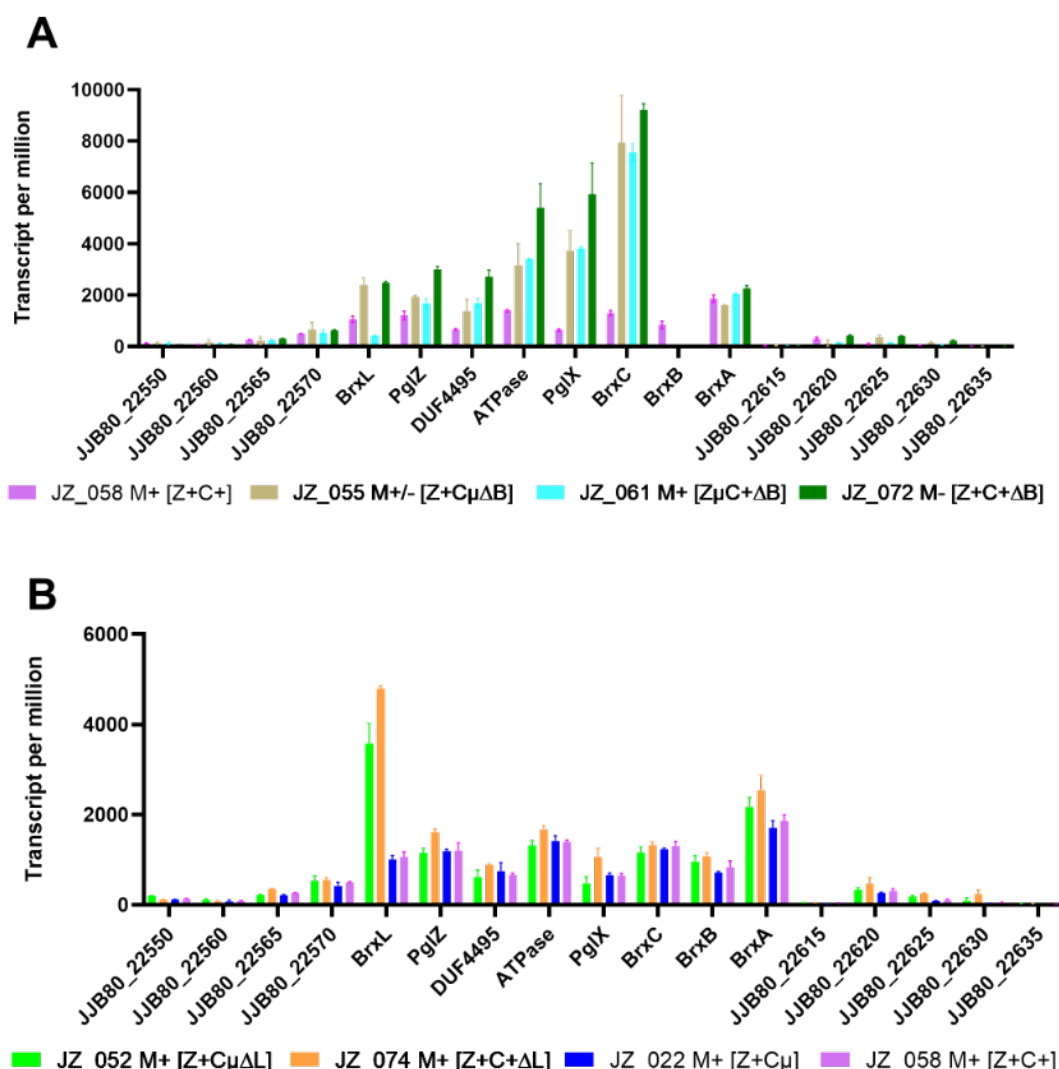
363 Figure 5: StySA locus transcription with *ΔbrxC::cat* and *ΔbrxL::cat.* The abcissa displays features in the

364 chromosome region with StySA/BREX locus and flanking features, not to scale, in order of chromosomal

365 coordinate. Transcription from *cat* and all CDSs within StySA is right to left (opposite to Figures 1 and 2).

366 Ordinate is transcripts mapping to the feature per million. Panel A: Strains: brown, $Z^+C\mu$ *ΔbrxB::cat;* blue,

367 $Z\mu C^+$ *ΔbrxB::cat*; green, Z^+C^+ *ΔbrxB::cat*; violet, Z^+C^+ (wild type). A lower plateau covering *pglX* and ATPase

368 (STM4494) for each strain following *brxC* may result from action of the *brxC*-distal terminator. A similar

369 lower plateau covering DUF4495 (STM4493) and *pglZ* suggests additional termination following ATPase,

370 for unknown reason. Lower transcription of *brxL* in JZ_061 is consistent with polarity seen in Figure 3

371 panel A. Panel B: In *ΔbrxL::cat,* the remnant 807 nt segment of *brxL* shows very high transcription from

372 the *cat* promoter in *ΔbrxL::cat* relative to isogenic *brxL⁺*. This transcription does not leak through the

- 19 -

373 bidirectional terminator into flanking features, e.g. JJB80_22570. Strain pairs: $Z^+C\mu$ background--green

374 $\Delta brxL::cat$, blue $brxL^+$; Z^+C^+ background--orange $\Delta brxL::cat$, violet $brxL^{+.}$. There may also be a small

375 increase in transcription of $brxA$-$brxB$ in the $\Delta brxL::cat$ strains.

Local and global inferences from transcriptome profiling

377 <u>$pglZ\mu$ is polar on $brxL$</u>

378 RNASeq measurements of transcription of the StySA locus (Figure 3 panel A)

379 suggest that the alternative alleles (μ or WT) do not affect transcription within the BREX

380 cassette, with one exception: transcription of $brxL$ is significantly increased when $pglZ$ is

381 wild type (JZ_058 and JZ_022) relative to $pglZ\mu$ (ER3625 and JZ_028). We infer that

382 translation termination at the $pglZ$ stop codon is polar on transcription of $brxL$. The

383 phenomenon of translational polarity is well known though the mechanism is still under

384 study (e.g., [55, 56]); transcript quantity is decreased following a translation stop.

385 However, the R$^-$ phenotype is not reversed simply by relief of polarity and increased

386 BrxL expression; JZ_022 carrying $pglZ^+$ $brxC\mu$ does not restrict (S3 Fig.). Intact BrxC is

387 also needed.

388 <u>Overexpression of 'brxL yields global effects without leakage across the insulating</u>

389 <u>terminator</u>

390 RNAseq results for $\Delta brxL::cat$ strains provide strong evidence that the left-hand

391 terminator insulates the flanking sequence from readthrough (Figure 5B). Inside the

392 locus, only the remnant '$brxL$ gene has altered expression; this is strongly increased

393 relative to isogenic $brxL^+$.

394 There are other effects of this disruption design. Growth rate is impaired for

395 $\Delta brxL::cat$ JZ_074 ($brxC^+$) though not JZ_052 ($brxC\mu$) (Figure 4B bottom panel); for

396   both strains, structural genes for two prophages are overexpressed, as well as numerous

397   other genes outside the locus (Figure 6, S5 File, S10 File).

398

399
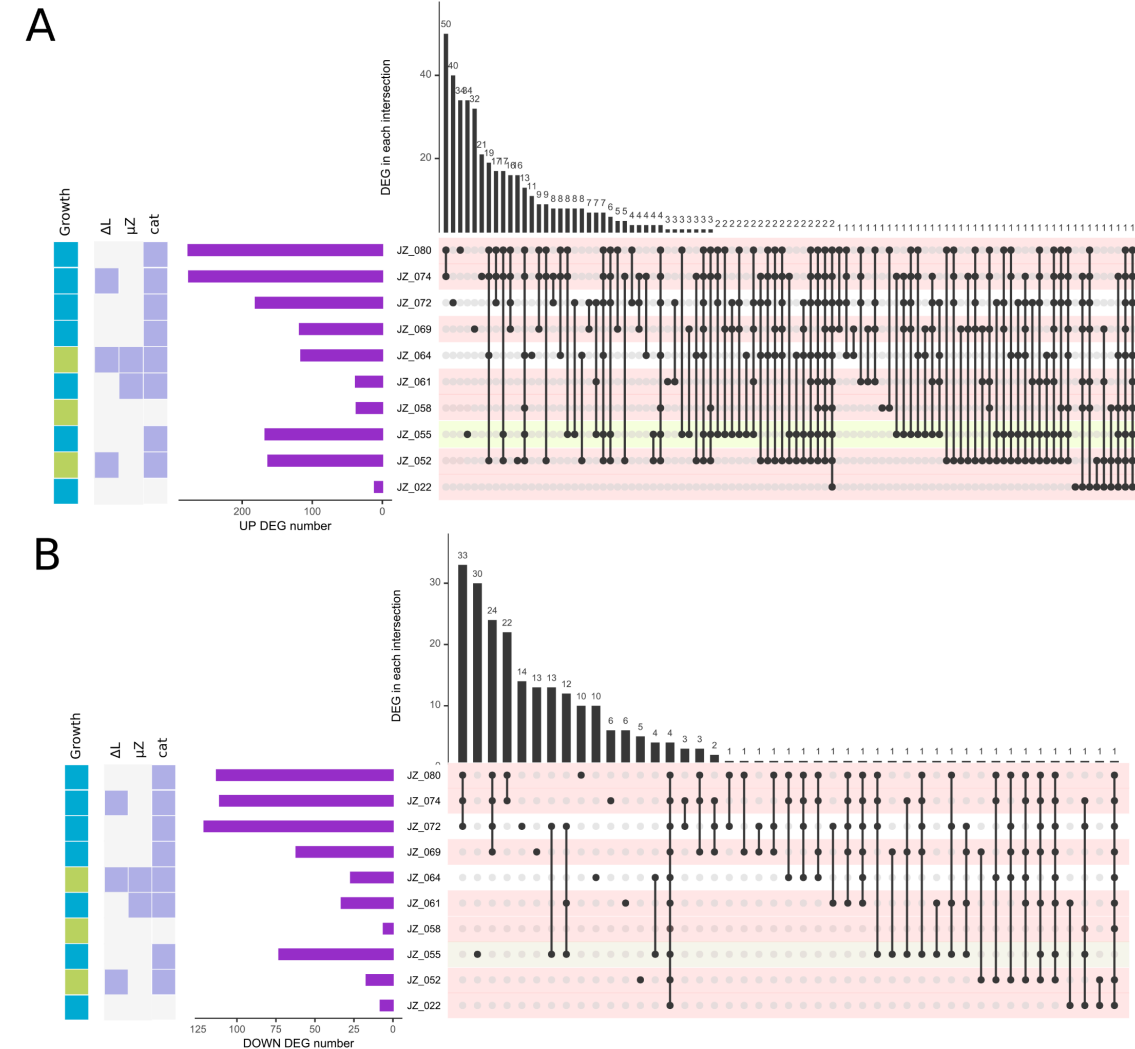
400   Figure 6 : Global changes in transcription visualized with UpsetR. Transcription for each CDS feature in

401   each strain was compared with ER3625, with differentially expressed genes (DEG) reported as adjusted

402   pvalue. On the right, sets of genes with changed expression and the sharing of sets with other strains are

403   represented by dots connected by lines. In each row, a black dot represents a set of genes in a particular

404   strain. Between rows, lines connect that dot (set) to other strains with dots representing the same set.

- 21 -

405　　The histogram at the top indicates the number of genes represented in that set (intersecting group). Top

406　　and bottom panels are respectively upregulated and downregulated sets of genes. In each row, levels of

407　　methylation are coded pink (M+), yellow (Mpartial) and white (M-). On the left, the purple bar chart

408　　indicates the total number of DEG in each strain. In the far left column, each strain was assigned a growth

409　　property, blue for slow growth, green for normal growth. In the middle columns, selected strain genotype

410　　information is color coded: white, marker is absent; blue, marker is present. Markers are: ΔL, the *cat*

411　　cassette has replaced the 5' segment of *brxL*; μZ, multiple mutations are present in *brxZ*; cat, the

412　　chloramphenicol cassette and its promoter are present. JZ_028 *(brxC⁺ pglZμ)* does not appear in the figure

413　　because it shows no change from ER3625 in phenotype or gene expression. The *cat* cassette is present in

414　　*brxB*: JZ_055, JZ_061, JZ_072; STM4493 (DUF4495): JZ_080 and in STM4494 (ATPase): JZ_069.

415　　<u>Unbalanced expression of BREX locus components with Δ*brxB::cat*</u>

416　　　　　The modification properties of strains with deletion/replacement of *brxB* are

417　　contradictory. In the genetic context of unmethylated strain $brxC^+$ *pglZμ* (JZ_028), the

418　　*cat* replacement of *brxB* (JZ_061) restores methylation (Figure 4 panel A, middle graph).

419　　In contrast, with modified strain *brxCμ* $pglZ^+$ (JZ_022), *cat* replacement of *brxB* in

420　　JZ_055 leads to a partial loss of methylation. In wild type $brxC^+$ $pglZ^+$(JZ_058) it leads

421　　to complete loss of methylation (JZ_072). In Discussion, we suggest that component

422　　ratios may account for this.

423　　　　　The "partially methylated" phenotype was reproducibly found in the genotype

424　　*brxCμ* $pglZ^+$Δ*brxB::cat*. Three independent constructions (JZ_055, JZ_056 and JZ_057)

425　　with the same configuration all display the same modification pattern. Dam sites are

426　　methylated normally. Examining the distribution of modified StySA sites among these

427　　isogenic strains, we observed in each a patchy modification pattern. Multiple nearby

428   StySA sites have very low modification while others comprise fully-modified regions,

429   rather than consistent half-modification (S6 File).

430   These RNAseq results for *ΔbrxB::cat* strains also support effective action of the

431   terminator between *brxC* and *pglX*: the transcript drops by half between these genes

432   (Figure 5A). This is consistent with the 5'RACE evidence for effective transcription

433   termination at the end of *brxC* (see above "Transcription overview" and S9 File).

434   Termination is not complete, allowing readthrough to *pglX*.

435   <u>Global inferences from transcriptome differential expression profiling</u>

436   The three engineered strains lacking *cat* cassettes are least globally affected in

437   terms of gene expression (Figure 6). Interestingly, JZ_028 ($brxC^+$ $pglZμ$) with unchanged

438   R^-M^- phenotype, has no significant differences with the ancestor (and does not appear in

439   Figure 6). JZ_022 ($brxCμ$ $pglZ^+$) has recovered modification (M^+); outside the locus, 11

440   genes show increased expression, 8 decreased. JZ_058 ($brxC^+$ $pglZ^+$) has recovered both

441   M^+ and restriction (R^+) with 37 changes up, 6 changes down. Of the 11 genes with

442   increased expression in M^+ JZ_022, 2 are unique to it: a possible operon of genes

443   JJB80_02810 (PLP-dependent aminotransferase) and JJB80_02815 (M15 family

444   metallopeptidase). Of the 37 genes increased in R^+M^+ (JZ_058), 1 is unique:

445   JJB80_02375, a ribosomal protein associated with changes in frameshifting. It seems

446   unlikely that these extra-locus changes. are responsible for the M and R phenotypes.

447   A very striking result is that strains with *cat* cassettes have very large numbers of

448   genes with significant changes in expression, and the sets of genes with altered

449   expression are not shared (Figure 6). We think it unlikely that the chloramphenicol

450   acetyltransferase enzyme itself is responsible; no drug was present. If these effects were

451   due to the action of the Cat protein (chloramphenicol transacetylase) without substrate,

452   all the *cat*-containing strains should have similar expression impact. However, both

453   number of genes and the shared sets are quite different. For example, with *cat* cassette in

454   *brxB*, the *pglZμ* strain JZ_061 has a fairly small number of genes affected--a number

455   similar to the number changed when wild type (*brxC⁺ pglZ⁺*) is restored, with no cassette

456   (JZ_058). The three Δ*brxL::cat* strains have a much bigger effect, which is even

457   enhanced by the presence of WT alleles of Z and C.

458   Instead it is likely that unbalanced expression of elements of the StySA island

459   mediate these drastic global effects. In most cases six, three or two downstream genes are

460   overexpressed. Possible action on off-target substrates by helicase, phosphatase and

461   ATPase activities is greatest for *brxB* (JZ_055, JZ_061, JZ_072) with overexpression of

462   *brxC, pglX*, STM4493-DUF4495, STM4494-ATPase, *pglZ* and *brxL*. For STM4493-

463   DUF4495 (JZ_080) downstream genes are STM4494-ATPase, *pglZ, brxL*. For

464   STM4494-ATPase (JZ_069), only *pglZ* and *brxL* are overexpressed. In general flavor,

465   these global changes affect the three prophages in the strain and cell surface composition

466   (S5 File).

467   Global changes in the Δ*brxL::cat* strains JZ_064, JZ_072 and JZ_074 are more

468   interesting, because the local effect of the *cat* insertion is limited: only the fragment of

469   the *brxL* gene, '*brxL*, is overexpressed (Figure 5B). Genes lying outside the StySA cluster

470   are insulated from transcription emanating from within by a very strong terminator. The

471   strong effects on global transcription (Figure 6 and S5 File) suggest that the '*brxL*

472   transcript or a translation product act outside the locus.

473   We favor the model that this partial gene is translated into a stand-alone protein

474   embodying the C-terminal domain (Lon_C, SSF54211: "Ribosomal_S5_D2-typ_fold";

475   see S7 File). The 5' 1278 nt segment of the gene has been replaced with the *cat* cassette,

476   but the 3' 807 nt still carries an ORF. Translation of the original *brxL* CDS terminates

477   within the added *cat* cassette (118 nt + 10 nt cassette). A transcript from the *cat* promoter

478   could allow translation restart following the *cat* CDS, at a CTG 20 nt into the *brxL*

479   remnant (52 nt from the *cat* UGA). Such translation would yield a protein carrying the

480   signature motifs from three domain annotation sources: GENE3d:3:30

481   Ribosomal_S5_D2-typ_fold_subgr; PFAM: Lon_C; Superfamily SSF54211

482   (Ribosomal_S5_D2-typ_fold). The candidate activities of the C-terminal domain include

483   hydrolysis, phosphoryl transfer and folding, any of which could explain major cellular

484   effects.

485   **<u>Discussion</u>**

486   <u>Genetic states:</u> Not all deletion derivatives designed were successfully constructed. Some

487   reasons for this are addressed here. Failure to recover deletion designs can be interpreted

488   two ways: overexpression of downstream genes in the cassette-containing intermediate

489   strain may be toxic; or the target protein itself is required for protection from toxicity of

490   other components of the locus.

491   Toxin/antitoxin pairs are frequently associated with defense islands [8]. Removal

492   of the antitoxin alone is usually lethal in such cases, while removal of the toxin alone is

493   tolerated. There are two possible such pairs within the StySA locus: BrxA-BrxB and

494   STM4493-STM4494.

495   Our failure to construct Δ*brxA* in three genetic contexts might be expected if BrxB

496   were a toxin and BrxA an antitoxin. However secondary transcription effects are also

497   anticipated (e.g. results of Δ*brxB::cat*), so the inference is weak.

498   STM4493-STM4494: We find that both genes can be disrupted individually in

499   three allelic contexts (Figure 4; S2 File). This argues against action as a toxin/antitoxin

500   pair, at least in this genetic environment.

501   We were also unable to recover a deletion of *pglX* in three allelic contexts, even

502   with a parent strain already lacking modification. Potentially the intermediate would

503   express STM4493-4494 aberrantly, rescuing a potential toxic effect not found otherwise.

504   We have not evaluated potential expression of a truncated PglX protein.

505   <u>Components of methyltransferase</u>

506   The N-terminal domain of BrxC and the complete PglZ protein are required for

507   DNA modification (Figure 3 panels B and C), while the C-terminus of BrxC is also

508   required for restriction (S3 Fig.). Modification and restriction are restored without

509   notable effects on transcription within the locus, except for relief of polarity on *brxL*

510   (Figure 3 panel A).

511   In agreement with findings for a related system from *E. coli* HS [34], BrxL is

512   required for restriction but not methylation. Accessory elements STM4494 and STM4493

513   are not essential for methylation.

514   BrxB is also not required for methylation activity and is unlikely to participate in

515   the active complex, though it could act as a chaperone or to support subcellular

516   localization. For example, overproduction of one component of transposase can result in

517   autoinhibition[57]; with multicomponent replication complexes, dominant-negative

518   phenotypes can result from formation of unproductive complexes or failure to remodel

519   them[58].

520   The contradictory effects of the Δ*brxB::cat* allele in the three genetic contexts

521   (Figure 4 panel A middle) is likely to relate to the mutational status of the *brxC* and *pglZ*

522   proteins overproduced (Figure 5A), rather than the expression level of StySA components

523   per se. The effect on StySA-BREX locus transcription is similar for all three strains:

524   transcription resulting from the strong *cat* promoter is 10-fold higher than in the wild-

525   type state for *brxC,* declining to ~twice the wild type in the *brxL* gene. Once again, the

526   effect on *brxL* transcription of the stop codon in *pglZμ* is seen.

527   <u>Overexpression of a BrxL fragment leads to action outside the StySA locus</u>

528   The genome-wide transcription effects of Δ*brxL::cat* are seen in three genetic

529   contexts, providing strong evidence for function of the truncated gene (Figure 6, S5 File).

530   We propose that this allele provides local opportunity for translation of a truncated

531   protein domain (S7 File). Such a solo domain might well escape control on its action

532   normally mediated by association with the other proteins in the BREX cluster. (see

533   "Global inferences from transcriptome profiling" and S5 File and S10 File). The intact

534   BrxL protein is critical to restriction action against phage L here (S3 File), in agreement

535   with [34]. A striking aspect of these global effects is enrichment for expression of

536   prophage genes, particularly Fels-1 and Gifsy-1 (S5 File). Regulation of expression in

537   these three prophages is tightly intertwined, mediated by antirepressor interactions [59]

538   and related to expression of pathogenicity functions [60]. Interestingly, the replication

539   regions of these phages are not overexpressed, while components of the virion are--the

- 27 -

540   phages are not induced to replicate, as would occur with DNA-damage induced escape.

541   This activity may shed light on the phage replication-inhibiting mode of action the role of

542   the intact restricting complex BrxL.

543   **Material and methods**

544   **Terminology**

545   ● StySA: genes and phenotypes related to the cluster of genes that specify REBASE

546   system *SenLT2II* of organism number 18099 and its lineal descendants [39]. This

547   cluster is characterized genetically in this work, using ER3625 as experimental

548   system. This segment of the ER3625 genome is descended from LT2 with

549   mutagenesis, although two distant genome segments of ER3625 descend from

550   *Salmonella enterica* serovar Abony SW803 [39].

551   ● RM: restriction-modification system

552   ● R^+, R^-: restriction phenotype, measured by reduction in plaque formation of

553   bacteriophage L when grown on a StySA M^- host [61]

554   ● M^+, M^-, $M^{+/-}$: StySA sites (GATCAG) are methylated at the second A fully, not at

555   all or partially.

556   ● *gene::Δcat*: a gene with a portion deleted and replaced with the chloramphenicol

557   resistance cassette of pKD3, including a strong promoter.

558   ● *brxC*, BrxC: gene, protein specified by LT2 locus_ID STM4496

559   ● *brxCμ,* BrxCμ: gene, protein specified by ER3625 locus_ID JJB80_ 22600

560   ● *pglZ,* PglZ: gene, protein specified by LT2 locus_ID STM4492

561   ● *pglZμ,* PglZμ: gene, protein specified by ER3625 locus_ID JJB80_ 22580

562   The rest of the DNA sequence of the ER3625 BREX cluster is identical in sequence to

563   that of LT2, so the genes and proteins will sometimes be referred to using the LT2

564   nomenclature. This will allow more ready access to database information since the LT2

565   sequence is well curated.

566   ● *brxA*: STM4498; JJB80_22610

567   ● *brxB*: STM4497; JJB80_22605

568   ● *pglX*: STM4495; JJB80_22595

569   ● STM4494: JJB80_22590, ATPase

570   ● STM4493: JJB80_22585, DUF4495

571   ● *brxL*: STM4491; JJB80_22575

572   ● '*brxL,* 'BrxL: N-terminally truncated gene and potential product resulting from

573   replacement of the 5' end of *brxL* with the *cat* cassette of pKD3, as described in

574   S7 File.

575   **Genome engineering**

576   The Datsenko and Wanner method [62] was adapted and combined with the

577   FAST-GE method [63] to engineer ER3625 descendants (Figure 7). During the second

578   step of classic λ Red engineering, drug resistance cassettes would be removed by

579   FLP/FRT recombination. For this, PCR-amplified *cat* cassette must be flanked by FRT

580   (FLP Recombinase Target) sites such that the Flp recombinase can drive specific

581   recombination between FRT sites. We found that resident FRT sites in ER3625

582   (*aroA::FRT, mrr::FRT::kan*) interfered with cassette addition to new sites, with the

583   newly synthesized fragment recombining preferentially into the resident sites. Thus, in

584   this work, the *cat* cassette was amplified and inserted without FRT sites.
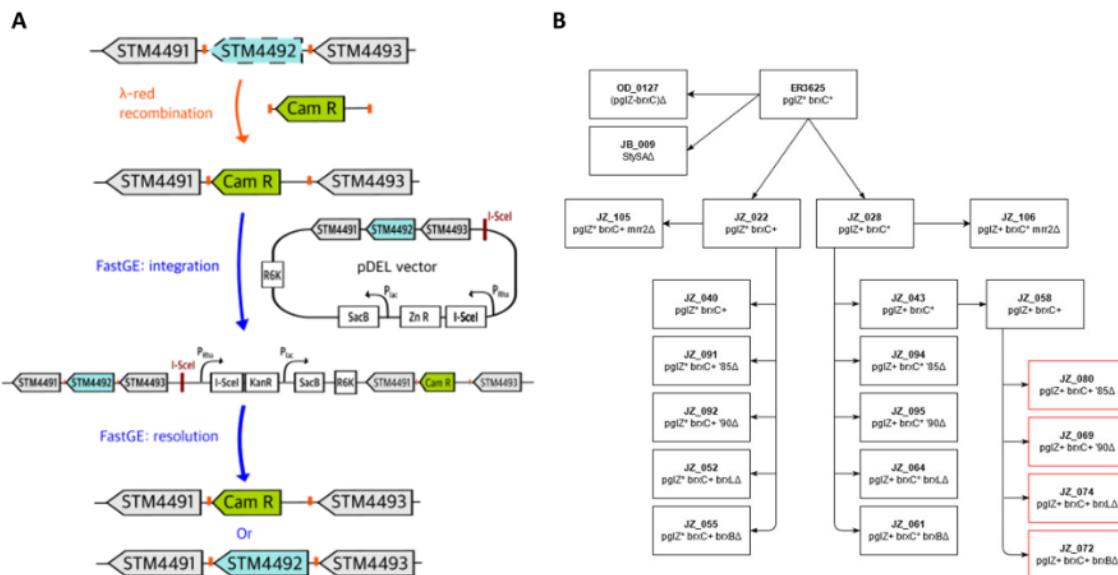
585

586 Figure 7: Engineering method and engineered strains. Panel A, Scarless engineering pipeline for

587 replacement of mutated gene by wild type. Panel B, partial strain pedigree. Intermediate steps of panel A

588 are omitted for clarity.

589 <u>Cassette replacement of a gene deletion or gene segment</u>

590   A *cat* cassette flanked by 36-50 bp of homology to the target gene was PCR-

591 amplified from pKD3 and used to electroporate an intermediate host carrying pKD46, a λ

592 Red- expressing thermosensitive vector [62]. Colony PCR with flanking gene-specific

593 primers validated the construction (S3 File).

594 <u>Replacement of the *cat* cassette with wild type sequence</u>

595   The FAST-GE method was developed to rapidly engineer the genomes of both *E.*

596 *coli* B and K12 strains with high efficiency by homologous recombination, with no

597 residual extra sequence ("scar") at the site of engineering. A single allele-exchange

598 vector pDEL, was designed as compact as possible to maximize flexibility in application

- 30 -

599   yield and fidelity. The pDEL vector contains several components involved in the

600   promotion of recombination processes such as I-SceI endonuclease with its

601   corresponding recognition site, and counterselection marker *sacB*. The unique double-

602   strand break caused by I-SceI was shown to improve local recombination efficiency.

603   Expression of *sacB* in the presence of sucrose is toxic to a variety of Gram-negative

604   bacteria. To improve the overall efficiency of the protocol, *sacB* and gene coding for I-

605   SceI endonuclease were placed under a *lac* and *rhaBAD* promoter respectively.

606   The original pDEL vector was constructed with a kanamycin resistance cassette

607   as the primary selection marker. Since our host carries a kanamycin resistance cassette at

608   the *mrr* locus, we designed two pDEL vectors with alternative antibiotic selection

609   markers: zeocin and gentamycin (see sequences in S4 File). Both drug resistance

610   cassettes were provided by Dr. Weigele. Zeocin is a bleomycin-like compound that kills

611   both eukaryotic and prokaryotic cells by introducing lethal double-strand breaks in

612   chromosomal DNA [64]. NEBuilder Hifi technology was used for construction of the

613   pDEL-Zn (pOD003) and pDEL-Gm (pOD004) vectors (S2 File). Briefly, zeocin and

614   gentamycin resistance cassettes were amplified by PCR with primers (oOD_070 and

615   oOD_071) containing extensions homologous to the linearized pDEL vector without its

616   kanamycin cassette (previously amplified with oOD_068 and oOD_069, S3 File), then

617   assembled with the vector, transformed into *E. coli pir+* cells, selecting for the new drug

618   resistance. Colony PCR with the set of primer oDO_072 and oDO_073 (S3 File) yields a

619   product of 988 bp or 1.1 kb for pOD003 or pOD004 vector respectively. An amplicon of

620   48 bp is expected for the original pDEL vector (S4 File and S3 File).

621   This combination of methods was used successfully to engineer the strains

622   characterized in this study for methylation and restriction activity (Figure 1B and S2

623   File). Cappable-seq data (Figure 2) were used in the strain design to limit as much as

624   possible the disruption of TSS and transcription terminator structures (S2 Fig.).

625   Linear method for locus replacement

626   This method consists of amplifying a cassette and flanking homology regions to

627   replace a genome segment with the cassette. Here we tested several lengths of homology

628   regions (3, 5, 9 and 12 kb). 3 kb worked. First, we amplified the Zn cassette from

629   pJBJZ_006 with oJB_018 and oJB_017 and the genomic flanking regions with dedicated

630   primers (oJB_009/oJB_006 and oJB_007/oJB_010). These fragments were purified with

631   NEB T1030 kit and ligated together using the NEBuilder® Hifi DNA assembly with a ratio

632   1:1:1 (HR1 : cassette : HR2) with a 30 min incubation. The assembly was then amplified

633   using Q5® High-Fidelity DNA Polymerase with an internal pair of primers

634   (oJB_005/oJB_008) and repurified with NEB T1030 kit. Transformation was performed

635   with 220 ng of the amplicons with Zn selection. Clones were verified by colony PCR.

636   **Phenotype tests**

637   Growth media

638   Bacteria were grown in RB (10g soy peptone, 5g yeast extract, 5g NaCl per liter) or

639   RBStrepKan (RB with Streptomycin (100 µg/mL) and Kanamycin (40 µg/mL) unless

640   otherwise indicated.

641   Phage restriction tests

642   Bacteriophage L, with 13 sites for StySA [61] was used to test for StySA

643 restriction, following the practice of the Colson and Ryu laboratories [40, 65]. Bacterial

644 strains were grown in RB with antibiotic overnight at 37°C with agitation. The cultures

645 were subcultured in new RB media without antibiotic and grown until exponential phase

646 at 37°C.

647   Two-layer agar plates were used for the spot titer restriction test (S3 Fig.). The

648 bottom layer is 1.5% agar (per liter: 15g of Bacto Agar BD Biosciences #214030, 10g of

649 Bacto Tryptone BD Biosciences #211699, 5g of Bacto Yeast extract BD Biosciences

650 #212720 and 5g of NaCl) and the top layer is an agar 0.7% (same recipe as the bottom

651 layer but only 7 g of Bacto agar). Bacterial cultures were mixed with the top agar layer

652 (56C) and poured on the bottom layer. The bacteriophage stocks (PH_JZ003 and

653 PH_JZ006, see details in S2 File) were diluted from 10^{-1} to 10^{-8}; 5 µl of each dilution was

654 spotted on the plates; incubated at room temperature until dry, and incubated 18h at 37°C.

655 Strains ER3625 and ER3649 were negative and positive controls. Plaques were counted

656 on spots where they were well isolated.

657 <u>Growth rate analysis</u>

658 Growth rates were measured in 96 well plates with a plate reader (REF) with two

659 technical replicates of each of three biological replicates. A single colony was inoculated

660 in 1 ml RBStrepKan in a deep well plate, then incubated overnight at 37°C with shaking

661 200 rpm. A 96 well plate (Greiner Bio-One ref: 655892) was prepared with 200µl

662 RBStrepKan per well and inoculated with 2µl of the overnight culture. The growth was

663 monitored every 15 minutes for 15 hours, between each measure, the plate was shaken.

664   Growth rates were calculated from the raw data using a rolling regression from:

665   https://padpadpadpad.github.io/post/calculating-microbial-growth-rates-from-od-using-

666   rolling-regression/.https://padpadpadpad.github.io/post/calculating-microbial-growth-

667   rates-from-od-using-rolling-regression/. A one-way ANOVA for normally distributed

668   samples of non-equal variance was performed on the data to determine statistical

669   significance of growth differences.

670   **Nucleic acid methods**

671   Genomic DNA (gDNA) extraction, sequencing

672   Each strain was growth in RB with the appropriate antibiotics (see "Growth

673   media" above) overnight at 37°C with 250 rpm agitation.

674   gDNA was extracted with the Monarch Genomic DNA purification kit (New

675   England Biolabs; Ipswich, MA, USA) from 1ml of culture.

676   Libraries from these genomic DNAs were sequenced using the PacBio RSII or

677   Sequel I sequencing platform. Briefly for RSII, SMRTbell libraries were constructed

678   from genomic DNA samples sheared to between 10 and 20 kb using the G-tubes protocol

679   (Covaris; Woburn, MA, USA), end repaired, and ligated to PacBio hairpin adapters.

680   Incompletely formed SMRTbell templates and linear DNAs were digested with a

681   combination of Exonuclease III and Exonuclease VII (New England Biolabs; Ipswich,

682   MA, USA). The SMRTbell library was prepared according to PacBio sample preparation

683   protocol sequenced with C4-P6 chemistry with a 300 min collection time.

684   For Sequel I libraries, SMRTbell libraries were constructed from genomic DNA

685   samples following the PacBio protocol for Sequel using the kit 100-938-900. DNA

686   qualification and quantification were performed using the Qubit fluorimeter (Invitrogen,

687   Eugene, OR) and 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA). The libraries

688   were prepared for binding following the PacBio guidelines generated by SMRT Link and

689   run on a Sequel I machine.

690   RNA extraction

691   For preculture three isolated colonies were grown in a 1ml RBStrepKan or

692   RBStrepKan with Chloramphenicol (30 µg/mL) overnight at 37°C in a deep well with

693   breathable cover tape. The cultures were subcultured the next day in 25ml RBStrepKan

694   (no chloramphenicol) at 37°C with 250 rpm agitation. Cells were harvested when OD 600

695   nm reached ~0.3 by centrifugation 10 min at 4°C. Pellets were resuspended in 100µl of

696   cold 0.1X PBS.

697   RNA was extracted using the Qiagen RNA extraction kit following the classical

698   protocol for bacterial RNA. The eluted RNA was then treated with DNAse I (NEB) and

699   then cleaned and concentrated using a classical phenol-chloroform RNA extraction. RNA

700   was stored at -80°C.

701   TSS determination using Cappable-seq method

702   The TSS Cappable-seq libraries were prepared following the recommendation of the

703   protocol from reference [47] starting from 2µg of RNA in duplicates and controls. The

704   libraries were run with a MiSeq with 1 X 75bp insert size using V3 Illumina platform.

705   The analysis was run in command line from the raw data using the script available at

706   https://github.com/Ettwiller/TSS/.

707   Quantitation of *brxX* transcription employed qPCR. Primers used are listed in S3

708   File. The Lunascript RT Supermix kit E3010 was used for random conversion of 500 ng

709   of extracted RNA to cDNA per samples (three biological replicates per strain) following

710   the kits guidelines. The no-RT reactions were run on the same plate with the same RNAs.

711   qPCR was then run with 2 primer sets, *yceB* (oJZ_116 and oJZ_117) and *brxX* (oJZ_150

712   and oJZ_151) using the Luna Universal qPCR Master Mix Protocol (M3003) with 1µl of

713   cDNA per well. For each primer pair a standard curve was run on the same plate as the

714   sample with 1, 0.1, 0.01, 0.001 and 0.0001ng and water only. All combinations were run

715   in duplicate on the same plate. The plates were sealed and centrifuged for 30s and then

716   run 39 cycles on the same CFX96 Touch Bio-Rad machine following the Luna kit

717   recommended cycle steps.

718   **Bioinformatic methods**

719   Annotation of predicted functional domains and transcription signals

720   Predicted functional protein domains were annotated using Genbank-assigned

721   protein IDs listed in S1 File for LT2 and for ER3625. The NCBI protein IDs are

722   automatically annotated with "regions" that correspond to Conserved Domain Database

723   [66] concise predictions. Additional automated domain annotations were generated and

724   visualized by submission to InterPro using the Geneious implementation of InterProScan

725   as described in S1 File. Manual search of the Conserved Domain Database with "Full

726   results" instead of "Concise results" also elicits annotations compatible with InterPro.

727   Potential transcription start sites were documented experimentally as described

728   below (CappableSeq). Rho-independent transcription terminators were predicted using

729   TransTerm HP algorithm version 2.09 ([67, 68]), then curated manually.

730 Sequence verification and methylation analysis

731   The BREX engineered locus sequence was verified in two ways. At first, PCR

732 were run with primers (oJZ_241/242, see S3 File) with LongAmp polymerase (New

733 England Biolabs; Ipswich, MA, USA), purified and sequenced (Sanger sequencing?). The

734 other method was to check the sequences of the modified locus with the PacBio

735 sequencing reads from the methylation analysis.

736 DNA motifs and degree of modification were generated using InterPulse Duration (IPD)

737 Ratios analyzed with RS_Modification_and_Motif_Analysis from PacBio as in [69, 70].

738 PBMotStat, a component of REBASE TOOLS [71] was used to calculate the % of

739 methylated sites with IPD > 2 for specific sites.

740   For the partially methylated strains, a .gff file of the methylation level of each

741 single base of the genome was downloaded from SMRT Link, filtered to keep only

742 significantly methylated sites, and uploaded in Geneious Prime on the reference genome

743 (File S6).

744 Global and local transcriptome analysis by RNAseq

745   RNAseq libraries were prepared with the "protocol for library preparation of

746 Intact RNA using NEBNext rRNA depletion kit (Bacteria) (NEB#E7850, NEB#78860)

747 and NEBNext Ultra II directional RNA library prep kit for Illumina (NEB#E7760,

748 NEB#E7765)" from 250ng of sample. Libraries were barcoded with dual index from

749 NEB Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs set 4, E6446S).

750 The sequencing of the pooled libraries was performed on a NextSeq apparatus

751   The raw reads were downloaded on Galaxy (Version 0.4.3.1) to perform the QC,

752   mapping, and feature assignment (S4 Fig.). After a first QC step to assess the quality of

753   the reads, they were trimmed using TrimGalore (Version 0.4.3) with parameters: -q 20 -s

754   3 -e 0.1 –length 20 and then a second QC steps was performed to verify the read quality

755   after trimming. The trimmed reads were mapped with bowtie2 (with default parameters

756   except –fr for upstream/downstream mate orientations) to the ER3625 genome (NCBI

757   CP067091-CP067092 and [61, 72]. The alignment was then used to count the number of

758   fragments mapped to CDS and rRNA features using in parallel Featurecounts [72]

759   (parameters: -s stranded (reverse)) and htseq (Version 0.9.1) with parameters: --mode

760   Union –stranded reverse –minaqual 10 --idattr locus_tag –nonunique no.

761   The downstream part of this analysis was performed using the Deseq2 R package

762   [73] in command line. The parameters were set as: lfcShrink = apeglm, padj < 0.05 and

763   log2FoldChange > |1.5|. UpsetR was used for visualization (Figure 6) [74]. Intersect (ref)

764   was used to generate lists of features common to chosen strain sets used in S5 File.

765 **<u>Acknowledgements</u>**

766 Eddie Gilcrease, Sherwood Casjens and Peter Weigele have been a wonderful source of

767 knowledge about the bacteriophage L and its use in *Salmonella* strains.

768 Bo Yan, Brad Langhorst, Zhiyi Sun and Amogh Changavi for bioinformatic advice and

769 training.

770 Julie Beaulieu and Jim Samuelson the pDEL vector and protocol and submission to

771 Addgene.

772   Barry Stoddard, Rick Morgan and Tim Blower for sharing ideas and discussions BREX

773   systems and their contexts.

774   New England Biolabs Research: Protein Expression and Modification Division and E.

775   coli club for discussions and "annoying" questions that helped build the project.

776   Donald Comb in memoriam for fostering basic research because it's exciting.

777   **<u>Supplementary Materials</u>**

778   S2_Fig_Sup_data_KO_segments.docx. The figure shows segments missing in *cat*::KO

779   constructions related to TSS and TTS; the document includes a legend

780   S3_Fig_colonies_restriction.docx. The figures show A. colony morphology for all

781   strains; B. Illustrative phage titer figure and table of restriction magnitude for all

782   strains. The document includes a legend.

783   S4_Fig_Galaxy_pipeline.docx. The document includes a legend with link to Galaxy

784   public web site.

785   S1_File_Fig_Table_protocol_Domain_predictions_six_StySA_components.docx. The

786   document includes a figure illustrating the domain hits identified, a table of

787   locus_tags and protein_IDs used for search, and a protocol used to generate the hit list

788   and visualization figure.

789   S2_File_plasmid_strain_phage.xlsx. Strain list with genotype and construction steps;

790   plasmids with description and availability; phage with source and identity and RM

791   phenotype of last host.

792   S3_File_Sup_Data_primers.xlsx. Oligonucleotide sequences used for plasmid

793   construction, strain engineering, validation of sequences and qPCR.

794   S4_File_ Sup_Data_Sequences_cassettes_plasmids.docx. This contains sequences

795   highlighted in color to navigate the locations of primers, coding sequences,

796   promoters, terminators, replicons, and protein binding sites.

797   S5_File_RNAseq_hits_annotated_and_coded.

798   S6_File_partial_methylation.zip. This file combines a reference sequence and gff file for

799   modification of independently constructed isogenic strains JZ_055, JZ_056, JZ_057.

800   S7_File_Potential_BrxL_activities.docx. This includes three figures and extended

801   discussion of activities of other proteins with domain hits like that of the BrxL C-

802   terminal domain.

803   S8_File_Sup_Data_SRA.xlsx. This describes the contents of SRA files: strain, NEB

804   strain ID if applicable, Biosample Name (NCBI-assigned), analysis type (RSII,

805   Sequel, Cappable-seq, NextSeq RNAseq)

806   S9_File_Transcript_Termination_Sites.docx. This document contains four figures:

807   strategy, and three IGV sequence read alignment screenshots at different

808   magnifications. No legend at present.

809   S10_File_ Protocol_for_adding_annotations_to_RNASeq_output_files

810   **Data availability (this is in a box at one side)**

811   All the data can be found in the BioProject PRJNA605961. All the raw data are

812   available under the SRA (See S8 File Sup_Data_SRA,xlsx).

813   All the Salmonella strains and phages are available from New England Biolabs.

814   The plasmids are available at Addgene.

815

816   REFERENCES

817

818   1.   Dimitriu T, Szczelkun MD, Westra ER. Evolutionary Ecology and Interplay of
819   Prokaryotic Innate and Adaptive Immune Systems. Curr Biol. 2020;30(19):R1189-R202.
820   Epub 2020/10/07. doi: 10.1016/j.cub.2020.08.028. PubMed PMID: 33022264; PubMed
821   Central PMCID: PMCPMC7116224.

822   2.   Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad
823   TF, Rohwer F, et al. Explaining microbial population genomics through phage predation.
824   Nat Rev Microbiol. 2009;7(11):828-36. Epub 2009/10/17. doi: 10.1038/nrmicro2235.
825   PubMed PMID: 19834481.

826   3.   Isaev AB, Musharova OS, Severinov KV. Microbial Arsenal of Antiviral
827   Defenses. Part II. Biochemistry (Mosc). 2021;86(4):449-70. Epub 2021/05/05. doi:
828   10.1134/S0006297921040064. PubMed PMID: 33941066.

829   4.   Isaev AB, Musharova OS, Severinov KV. Microbial Arsenal of Antiviral
830   Defenses – Part I. Biochemistry (Moscow). 2021;86(3):319-37. doi:
831   10.1134/s0006297921030081 PMID - 33838632.

832   5.   Azam AH, Tanji Y. Bacteriophage-host arm race: an update on the mechanism of
833   phage resistance in bacteria and revenge of the phage with the perspective for phage
834   therapy. Appl Microbiol Biotechnol. 2019;103(5):2121-31. Epub 2019/01/27. doi:
835   10.1007/s00253-019-09629-x. PubMed PMID: 30680434.

836   6.   Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their
837   phage foes. Nature. 2020;577(7790):327-36. doi: 10.1038/s41586-019-1894-8 PMID -
838   31942051.

839   7.   Gao L, Altae-Tran H, Bohning F, Makarova KS, Segel M, Schmid-Burgk JL, et
840   al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. Science.
841   2020;369(6507):1077-84. Epub 2020/08/29. doi: 10.1126/science.aba0372. PubMed
842   PMID: 32855333; PubMed Central PMCID: PMCPMC7985843.

843   8.   Koonin EV, Makarova KS, Wolf YI, Krupovic M. Evolutionary entanglement of
844   mobile genetic elements and host defence systems: guns for hire. Nat Rev Genet.

845 2020;21(2):119-31. Epub 2019/10/16. doi: 10.1038/s41576-019-0172-9. PubMed PMID:
846 31611667.

847 9.   Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A
848 nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases
849 and their genes. Nucleic Acids Res. 2003;31(7):1805-12. Epub 2003/03/26. doi:
850 10.1093/nar/gkg274. PubMed PMID: 12654995; PubMed Central PMCID:
851 PMCPMC152790.

852 10.   Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. Highlights of the
853 DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 2014;42(1):3-
854 19. Epub 2013/10/22. doi: 10.1093/nar/gkt990. PubMed PMID: 24141096; PubMed
855 Central PMCID: PMCPMC3874209.

856 11.   Wang L, Jiang S, Deng Z, Dedon PC, Chen S. DNA phosphorothioate
857 modification-a new multi-functional epigenetic system in bacteria. FEMS Microbiol Rev.
858 2019;43(2):109-22. Epub 2018/10/06. doi: 10.1093/femsre/fuy036. PubMed PMID:
859 30289455; PubMed Central PMCID: PMCPMC6435447.

860 12.   Oliveira PH, Fang G. Conserved DNA Methyltransferases: A Window into
861 Fundamental Mechanisms of Epigenetic Regulation in Bacteria. Trends in Microbiology.
862 2020;29(1):28-40. doi: 10.1016/j.tim.2020.04.007 PMID - 32417228.

863 13.   Anton BP, Roberts RJ. Beyond Restriction Modification: Epigenomic Roles of
864 DNA Methylation in Prokaryotes. Annu Rev Microbiol. 2021;75(1):129-49. Epub
865 2021/07/28. doi: 10.1146/annurev-micro-040521-035040. PubMed PMID: 34314594.

866 14.   Rifat D, Wright NT, Varney KM, Weber DJ, Black LW. Restriction endonuclease
867 inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. J Mol Biol.
868 2008;375(3):720-34. Epub 2007/11/27. doi: 10.1016/j.jmb.2007.10.064. PubMed PMID:
869 18037438; PubMed Central PMCID: PMCPMC2255585.

870 15.   Tock MR, Dryden DT. The biology of restriction and anti-restriction. Curr Opin
871 Microbiol. 2005;8(4):466-72. Epub 2005/06/28. doi: 10.1016/j.mib.2005.06.003.
872 PubMed PMID: 15979932.

873 16.   Isaev A, Drobiazko A, Sierro N, Gordeeva J, Yosef I, Qimron U, et al. Phage T7
874 DNA mimic protein Ocr is a potent inhibitor of BREX defence. Nucleic Acids Res.
875 2020;48(10):5397-406. Epub 2020/04/28. doi: 10.1093/nar/gkaa290. PubMed PMID:
876 32338761; PubMed Central PMCID: PMCPMC7261183.

877 17.   González-Montes L, del Campo I, Garcillán-Barcia MP, de la Cruz F, Moncalian
878 G. ArdC, a ssDNA-binding protein with a metalloprotease domain, overpasses the
879 recipient hsdRMS restriction system broadening conjugation host range. PLoS Genet.
880 2020;16(4):e1008750-22. doi: 10.1371/journal.pgen.1008750. PubMed PMID:
881 B37483D8-4993-43E1-98D5-14F102565561.

882 18.   Roberts GA, Stephanou AS, Kanwar N, Dawson A, Cooper LP, Chen K, et al.
883 Exploring the DNA mimicry of the Ocr protein of phage T7. Nucleic Acids Res.
884 2012;40(16):8129-43. Epub 2012/06/12. doi: 10.1093/nar/gks516. PubMed PMID:
885 22684506; PubMed Central PMCID: PMCPMC3439906.

886   19.   Ho C-H, Wang H-C, Ko T-P, Chang Y-C, Wang AH-J. The T4 phage DNA
887   mimic protein Arn inhibits the DNA binding activity of the bacterial histone-like protein
888   H-NS. Journal of Biological Chemistry. 2014;289(39):27046-54. doi:
889   10.1074/jbc.M114.590851. PubMed PMID: 25118281; PubMed Central PMCID:
890   PMCPMC4175342.

891   20.   Iida S, Hiestand-Nauer R, Sandmeier H, Lehnherr H, Arber W. Accessory genes
892   in the darA operon of bacteriophage P1 affect antirestriction function, generalized
893   transduction, head morphogenesis, and host cell lysis. Virology. 1998;251(1):49-58.
894   Epub 1998/11/14. doi: 10.1006/viro.1998.9405. PubMed PMID: 9813202.

895   21.   Makarova KS, Wolf YI, Koonin EV. Comparative genomics of defense systems
896   in archaea and bacteria. Nucleic Acids Res. 2013;41(8):4360-77. Epub 2013/03/09. doi:
897   10.1093/nar/gkt157. PubMed PMID: 23470997; PubMed Central PMCID:
898   PMCPMC3632139.

899   22.   Makarova KS, Wolf YI, Snir S, Koonin EV. Defense islands in bacterial and
900   archaeal genomes and prediction of novel defense systems. J Bacteriol.
901   2011;193(21):6039-56. Epub 2011/09/13. doi: 10.1128/JB.05535-11. PubMed PMID:
902   21908672; PubMed Central PMCID: PMCPMC3194920.

903   23.   Sibley MH, Raleigh EA. Cassette-like variation of restriction enzyme genes in
904   Escherichia coli C and relatives. Nucleic Acids Res. 2004;32(2):522-34. Epub
905   2004/01/28. doi: 10.1093/nar/gkh194. PubMed PMID: 14744977; PubMed Central
906   PMCID: PMCPMC373321.

907   24.   Raleigh EA. Organization and function of the mcrBC genes of Escherichia coli K-
908   12. Mol Microbiol. 1992;6(9):1079-86. Epub 1992/05/01. doi: 10.1111/j.1365-
909   2958.1992.tb01546.x. PubMed PMID: 1316984.

910   25.   Kelleher JE, Raleigh EA. A novel activity in Escherichia coli K-12 that directs
911   restriction of DNA modified at CG dinucleotides. J Bacteriol. 1991;173(16):5220-3.
912   Epub 1991/08/01. doi: 10.1128/jb.173.16.5220-5223.1991. PubMed PMID: 1830580;
913   PubMed Central PMCID: PMCPMC208216.

914   26.   Raleigh EA, Trimarchi R, Revel H. Genetic and physical mapping of the mcrA
915   (rglA) and mcrB (rglB) loci of Escherichia coli K-12. Genetics. 1989;122(2):279-96.
916   Epub 1989/06/01. doi: 10.1093/genetics/122.2.279. PubMed PMID: 2548920; PubMed
917   Central PMCID: PMCPMC1203701.

918   27.   Koonin EV, Zhang F. Coupling immunity and programmed cell suicide in
919   prokaryotes: Life-or-death choices. BioEssays. 2017;39(1):1-9. doi:
920   10.1002/bies.201600186 PMID - 27896818.

921   28.   Montgomery MT, Guerrero Bustamante CA, Dedrick RM, Jacobs-Sera D, Hatfull
922   GF. Yet More Evidence of Collusion: a New Viral Defense System Encoded by Gordonia
923   Phage CarolAnn. mBio. 2019;10(2). Epub 2019/03/21. doi: 10.1128/mBio.02417-18.
924   PubMed PMID: 30890601; PubMed Central PMCID: PMCPMC6426606.

925   29.   Hoskisson PA, Sumby P, Smith MCM. The phage growth limitation system in
926   Streptomyces coelicolor A(3)2 is a toxin/antitoxin system, comprising enzymes with
927   DNA methyltransferase, protein kinase and ATPase activity. Virology. 2015;477:100-9.

928   Epub 2015/01/17. doi: 10.1016/j.virol.2014.12.036. PubMed PMID: 25592393; PubMed
929   Central PMCID: PMCPMC4365076.

930   30.   Sumby P, Smith MC. Phase variation in the phage growth limitation system of
931   Streptomyces coelicolor A3(2). J Bacteriol. 2003;185(15):4558-63. Epub 2003/07/18.
932   doi: 10.1128/JB.185.15.4558-4563.2003. PubMed PMID: 12867465; PubMed Central
933   PMCID: PMCPMC165769.

934   31.   Sumby P, Smith MC. Genetics of the phage growth limitation (Pgl) system of
935   Streptomyces coelicolor A3(2). Mol Microbiol. 2002;44(2):489-500. Epub 2002/04/26.
936   doi: 10.1046/j.1365-2958.2002.02896.x. PubMed PMID: 11972785.

937   32.   Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, et
938   al. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J.
939   2015;34(2):169-83. Epub 2014/12/03. doi: 10.15252/embj.201489455. PubMed PMID:
940   25452498; PubMed Central PMCID: PMCPMC4337064.

941   33.   Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide
942   experimental determination of barriers to horizontal gene transfer. Science.
943   2007;318(5855):1449-52. Epub 2007/10/20. doi: 10.1126/science.1147112. PubMed
944   PMID: 17947550.

945   34.   Gordeeva J, Morozova N, Sierro N, Isaev A, Sinkunas T, Tsvetkova K, et al.
946   BREX system of Escherichia coli distinguishes self from non-self by methylation of a
947   specific DNA site. Nucleic Acids Res. 2019;47(1):253-65. Epub 2018/11/13. doi:
948   10.1093/nar/gky1125. PubMed PMID: 30418590; PubMed Central PMCID:
949   PMCPMC6326788.

950   35.   Hui W, Zhang W, Kwok LY, Zhang H, Kong J, Sun T. A Novel Bacteriophage
951   Exclusion (BREX) System Encoded by the pglX Gene in Lactobacillus casei Zhang.
952   Appl Environ Microbiol. 2019;85(20):6-12. Epub 2019/08/11. doi: 10.1128/AEM.01001-
953   19. PubMed PMID: 31399407; PubMed Central PMCID: PMCPMC6805096.

954   36.   Picton DM, Luyten YA, Morgan RD, Nelson A, Smith Darren L, Dryden DTF, et
955   al. The phage defence island of a multidrug resistant plasmid uses both BREX and type
956   IV restriction for complementary protection from viruses. Nucleic Acids Research. 2021.
957   doi: 10.1093/nar/gkab906 PMID - 34657954.

958   37.   Hattman S, Schlagman S, Goldstein L, Frohlich M. *Salmonella typhimurium* SA
959   host specificity system is based on deoxyribonucleic acid-adenine methylation. J
960   Bacteriol. 1976;127(1):211-7. Epub 1976/07/01. doi: 10.1128/jb.127.1.211-217.1976.
961   PubMed PMID: 776925; PubMed Central PMCID: PMCPMC233053.

962   38.   Bullas LR, Colson C, Neufeld B. Deoxyribonucleic acid restriction and
963   modification systems in Salmonella: chromosomally located systems of different
964   serotypes. Journal of Bacteriology. 1980;141(1):275-92. doi: 10.1128/jb.141.1.275-
965   292.1980 PMID - 6243623.

966   39.   Zaworski J, Dagva O, Kingston AW, Fomenkov A, Morgan RD, Bossi L, et al.
967   Genome archaeology of two laboratory Salmonella enterica enterica sv Typhimurium. G3
968   (Bethesda). 2021;11(9):jkab226-. Epub 2021/09/21. doi: 10.1093/g3journal/jkab226.
969   PubMed PMID: 34544129; PubMed Central PMCID: PMCPMC8496262.

970 40.   Bullas LR, Ryu JI. *Salmonella typhimurium* LT2 strains which are r- m+ for all
971 three chromosomally located systems of DNA restriction and modification. J Bacteriol.
972 1983;156(1):471-4. Epub 1983/10/01. doi: 10.1128/jb.156.1.471-474.1983. PubMed
973 PMID: 6352690; PubMed Central PMCID: PMCPMC215113.

974 41.   Bishop AL, Baker S, Jenks S, Fookes M, Gaora PO, Pickard D, et al. Analysis of
975 the hypervariable region of the Salmonella enterica genome associated with tRNA(leuX).
976 J Bacteriol. 2005;187(7):2469-82. Epub 2005/03/19. doi: 10.1128/JB.187.7.2469-
977 2482.2005. PubMed PMID: 15774890; PubMed Central PMCID: PMCPMC1065210.

978 42.   Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, et al.
979 RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res.
980 2018;46(D1):D851-D60. Epub 2017/11/08. doi: 10.1093/nar/gkx1068. PubMed PMID:
981 29112715; PubMed Central PMCID: PMCPMC5753331.

982 43.   McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et
983 al. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.
984 Nature. 2001;413(6858):852-6. Epub 2001/10/26. doi: 10.1038/35101614. PubMed
985 PMID: 11677609.

986 44.   Bullas LR, Colson C. DNA restriction and modification systems in Salmonella.
987 III. SP, a Salmonella potsdam system allelic to the SB system in Salmonella
988 typhimurium. Mol Gen Genet. 1975;139(3):177-88. Epub 1975/08/27. PubMed PMID:
989 1102934.

990 45.   Tsai SP, Hartin RJ, Ryu J. Transformation in restriction-deficient Salmonella
991 typhimurium LT2. J Gen Microbiol. 1989;135(9):2561-7. Epub 1989/09/01. doi:
992 10.1099/00221287-135-9-2561. PubMed PMID: 2697751.

993 46.   Haack KR, Roth JR. Recombination between chromosomal IS200 elements
994 supports frequent duplication formation in Salmonella typhimurium. Genetics.
995 1995;141(4):1245-52.

996 47.   Ettwiller L, Buswell J, Yigit E, Schildkraut I. A novel enrichment strategy reveals
997 unprecedented number of novel transcription start sites at single base resolution in a
998 model prokaryote and the gut microbiome. BMC Genomics. 2016;17(1):199. Epub
999 2016/03/10. doi: 10.1186/s12864-016-2539-z. PubMed PMID: 26951544; PubMed
1000 Central PMCID: PMCPMC4782308.

1001 48.   Kroger C, Colgan A, Srikumar S, Handler K, Sivasankaran SK, Hammarlof DL,
1002 et al. An infection-relevant transcriptomic compendium for Salmonella enterica Serovar
1003 Typhimurium. Cell Host Microbe. 2013;14(6):683-95. Epub 2013/12/18. doi:
1004 10.1016/j.chom.2013.11.010. PubMed PMID: 24331466.

1005 49.   Morgan RD, Dwinell EA, Bhatia TK, Lang EM, Luyten YA. The MmeI family:
1006 type II restriction-modification enzymes that employ single-strand modification for host
1007 protection. Nucleic Acids Res. 2009;37(15):5208-21. Epub 2009/07/07. doi:
1008 10.1093/nar/gkp534. PubMed PMID: 19578066; PubMed Central PMCID:
1009 PMCPMC2731913.

1010 50.   Furuta Y, Kobayashi I. Movement of DNA sequence recognition domains
1011 between non-orthologous proteins. Nucleic Acids Res. 2012;40(18):9218-32. Epub

1012 2012/07/24. doi: 10.1093/nar/gks681. PubMed PMID: 22821560; PubMed Central
1013 PMCID: PMCPMC3467074.

1014 51.   Bayliss CD, Callaghan MJ, Moxon ER. High allelic diversity in the
1015 methyltransferase gene of a phase variable type III restriction-modification system has
1016 implications for the fitness of Haemophilus influenzae. Nucleic Acids Res.
1017 2006;34(14):4046-59. Epub 2006/08/18. doi: 10.1093/nar/gkl568. PubMed PMID:
1018 16914439; PubMed Central PMCID: PMCPMC1557822.

1019 52.   Lauster R, Trautner TA, Noyer-Weidner M. Cytosine-specific type II DNA
1020 methyltransferases. A conserved enzyme core with variable target-recognizing domains. J
1021 Mol Biol. 1989;206(2):305-12. Epub 1989/03/20. doi: 10.1016/0022-2836(89)90480-4.
1022 PubMed PMID: 2716049.

1023 53.   Loenen WA, Dryden DT, Raleigh EA, Wilson GG. Type I restriction enzymes
1024 and their relatives. Nucleic Acids Res. 2014;42(1):20-44. Epub 2013/09/27. doi:
1025 10.1093/nar/gkt847. PubMed PMID: 24068554; PubMed Central PMCID:
1026 PMCPMC3874165.

1027 54.   Colson C, Colson AM. A new *Salmonella typhimurium* DNA host specificity. J
1028 Gen Microbiol. 1971;69(3):345-51. Epub 1971/12/01. doi: 10.1099/00221287-69-3-345.
1029 PubMed PMID: 4947313.

1030 55.   de Smit MH, Verlaan PW, van Duin J, Pleij CW. Intracistronic transcriptional
1031 polarity enhances translational repression: a new role for Rho. Mol Microbiol.
1032 2008;69(5):1278-89. Epub 2009/01/28. doi: 10.1111/j.1365-2958.2008.06360.x. PubMed
1033 PMID: 19172759.

1034 56.   Chen M, Fredrick K. Measures of single- versus multiple-round translation argue
1035 against a mechanism to ensure coupling of transcription and translation. Proc Natl Acad
1036 Sci USA. 2018;115(42):10774-9. Epub 2018/10/03. doi: 10.1073/pnas.1812940115.
1037 PubMed PMID: 30275301; PubMed Central PMCID: PMCPMC6196535.

1038 57.   Lohe AR, Hartl DL. Autoregulation of mariner transposase activity by
1039 overproduction and dominant-negative complementation. Molecular Biology and
1040 Evolution. 1996;13(4):549-55. doi: 10.1093/oxfordjournals.molbev.a025615 PMID -
1041 8882498.

1042 58.   Gaubitz C, Liu X, Magrino J, Stone NP, Landeck J, Hedglin M, et al. Structure of
1043 the human clamp loader reveals an autoinhibited conformation of a substrate-bound
1044 AAA+ switch. Proc Natl Acad Sci USA. 2020;117(38):23571-80. doi:
1045 10.1073/pnas.2007437117 PMID - 32907938.

1046 59.   Lemire S, Figueroa-Bossi N, Bossi L. Bacteriophage crosstalk: coordination of
1047 prophage induction by trans-acting antirepressors. PLoS Genet. 2011;7(6):e1002149.
1048 Epub 2011/07/07. doi: 10.1371/journal.pgen.1002149. PubMed PMID: 21731505;
1049 PubMed Central PMCID: PMCPMC3121763.

1050 60.   Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L. Variable assortment of
1051 prophages provides a transferable repertoire of pathogenic determinants in Salmonella.
1052 Mol Microbiol. 2001;39(2):260-71. Epub 2001/01/03. doi: 10.1046/j.1365-
1053 2958.2001.02234.x. PubMed PMID: 11136448.

61.   Zaworski J, McClung C, Ruse C, Weigele PR, Hendrix RW, Ko CC, et al. Genome analysis of Salmonella enterica serovar Typhimurium bacteriophage L, indicator for StySA (StyLT2III) restriction-modification system action. G3 (Bethesda). 2021;11(1). Epub 2021/02/10. doi: 10.1093/g3journal/jkaa037. PubMed PMID: 33561243; PubMed Central PMCID: PMCPMC8022706.

62.   Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA. 2000;97(12):6640-5. Epub 2000/06/01. doi: 10.1073/pnas.120163297. PubMed PMID: 10829079; PubMed Central PMCID: PMCPMC18686.

63.   Tikh IB, Samuelson JC. Leveraging modern DNA assembly techniques for rapid, markerless genome modification. Biol Methods Protoc. 2016;1(1):bpw004. Epub 2016/12/27. doi: 10.1093/biomethods/bpw004. PubMed PMID: 32368618; PubMed Central PMCID: PMCPMC7189271.

64.   Gatignol A, Durand H, Tiraby G. Bleomycin resistance conferred by a drug-binding protein. FEBS Lett. 1988;230(1-2):171-5. Epub 1988/03/28. doi: 10.1016/0014-5793(88)80665-3. PubMed PMID: 2450783.

65.   Colson C, Van Pel A. DNA restriction and modification systems in Salmonella. I. SA and SB, two Salmonella typhimurium systems determined by genes with a chromosomal location comparable to that of the Escherichia coli hsd genes. Mol Gen Genet. 1974;129(4):325-37. Epub 1974/04/03. doi: 10.1007/BF00265696. PubMed PMID: 4601252.

66.   Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research. 2019;48(D1):D265-D8. doi: 10.1093/nar/gkz991 PMID - 31777944.

67.   Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biology. 2007;8(2):R22. doi: 10.1186/gb-2007-8-2-r22 PMID - 17313685.

68.   Ermolaeva MD, Khalak HG, White O, Smith HO, Salzberg SL. Prediction of transcription terminators in bacterial genomes11Edited by F. E. Cohen. Journal of Molecular Biology. 2000;301(1):27-33. doi: 10.1006/jmbi.2000.3836 PMID - 10926490.

69.   Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. 2012;40(4):e29. Epub 2011/12/14. doi: 10.1093/nar/gkr1146. PubMed PMID: 22156058; PubMed Central PMCID: PMCPMC3287169.

70.   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;7(6):461-5. Epub 2010/05/11. doi: 10.1038/nmeth.1459. PubMed PMID: 20453866; PubMed Central PMCID: PMCPMC2879396.

71.   Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res.

1096   2015;43(Database issue):D298-9. Epub 2014/11/08. doi: 10.1093/nar/gku1046. PubMed
1097   PMID: 25378308; PubMed Central PMCID: PMCPMC4383893.

1098   72.   Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program
1099   for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.
1100   Epub 2013/11/15. doi: 10.1093/bioinformatics/btt656. PubMed PMID: 24227677.

1101   73.   Love MI, Huber W, Anders S. Moderated estimation of fold change and
1102   dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. Epub
1103   2014/12/18. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed
1104   Central PMCID: PMCPMC4302049.

1105   74.   Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of
1106   intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-40. doi:
1107   10.1093/bioinformatics/btx364 PMID - 28645171.

1108