Transposon-insertion sequencing screens unveil requirements for EHEC growth and intestinal colonization

Alyson R. Warr[1,2,#], Troy P. Hubbard[1,2,#], Diana Munera[1,2,3,#], Carlos J. Blondel[1,2,4,#], Pia Abel zur Wiesch[1,2,5], Sören Abel[1,2,5], Xiaoxue Wang[1,2,6], Brigid M Davis[1,2] & Matthew K. Waldor[1,2,7*]

Affiliations

[1]Division of Infectious Diseases, Brigham & Women's Hospital, Boston, MA, USA, 02115
[2]Department of Microbiology, Harvard Medical School, Boston, MA, USA, 02115
[3]DM current address: DNA Electronics, London, UK.
[4]CJB current address: Instituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile
[5]PAW and SA current address: Department of Pharmacy, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway and the Centre for Molecular Medicine Norway, Nordic EMBL Partnership, Oslo, Norway
[6]XW current address: Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China
[7]HHMI, Boston, MA, USA, 02115.
# These authors contributed equally to this work

*Correspondence to: Matthew Waldor, 181 Longwood Avenue, Boston, MA, 02115.
Phone: 617- 525- 4646, Fax: 617-525-4660, mwaldor@research.bwh.harvard.edu

Abstract

Enterohemorrhagic *Escherichia coli* O157:H7 (EHEC) is an important food-borne pathogen that colonizes the colon. Transposon-insertion sequencing (TIS) was used to identify genes required for EHEC and commensal *E. coli* K-12 growth in vitro and for EHEC growth in vivo in the infant rabbit colon. Surprisingly, many conserved loci contribute to EHEC's but not to K-12's growth in vitro, suggesting that gene acquisition during EHEC evolution has heightened the pathogen's reliance on certain metabolic processes that are dispensable for K-12. There was a restrictive bottleneck for EHEC colonization of the rabbit colon, which complicated identification of EHEC genes facilitating growth in vivo. Both a refined version of an existing analytic framework as well as PCA-based analysis were used to compensate for the effects of the infection bottleneck. These analyses confirmed that the EHEC LEE-encoded type III secretion apparatus is required for growth in vivo and revealed that only a few effectors are critical for in vivo fitness. Numerous mutants not previously associated with EHEC survival/growth in vivo also appeared attenuated in vivo, and a subset of these putative in vivo fitness factors were validated. Some were found to contribute to efficient type-three secretion while others, including *tatABC, oxyR*, *envC*, *acrAB*, and *cvpA,* promote EHEC resistance to host-derived stresses encountered in vivo. *cvpA*, which is also required for intestinal growth of several other enteric pathogens, proved to be required for EHEC, *Vibrio cholerae* and *Vibrio parahaemolyticus* resistance to the bile salt deoxycholate. Collectively, our findings provide a comprehensive framework for understanding EHEC growth in the intestine.

53 **Author Summary**

54 Enterohemorrhagic *E. coli* (EHEC) are important food-borne pathogens that infect the
55 colon. We created a highly saturated EHEC transposon library and used transposon
56 insertion sequencing to identify the genes required for EHEC growth in vitro and in vivo
57 in the infant rabbit colon. We found that there is a large infection bottleneck in the rabbit
58 model of intestinal colonization, and refined two analytic approaches to facilitate
59 rigorous identification of new EHEC genes that promote fitness in vivo. Besides the
60 known type III secretion system, more than 200 additional genes were found to
61 contribute to EHEC survival and/or growth within the intestine. The requirement for
62 some of these new in vivo fitness factors was confirmed, and their contributions to
63 infection were investigated. This set of genes should be of considerable value for future
64 studies elucidating the processes that enable the pathogen to proliferate in vivo and for
65 design of new therapeutics.

66

Introduction

Enterohemorrhagic Escherichia coli (EHEC) is an important food-borne pathogen that causes gastrointestinal (GI) infections worldwide. EHEC is a non-invasive pathogen that colonizes the human colon and gives rise to sporadic infections as well as large outbreaks (reviewed in (1–3)). The clinical consequences of EHEC infection range from mild diarrhea to hemorrhagic colitis and include the potentially lethal hemolytic uremic syndrome (HUS) (4,5).

The paradigmatic EHEC O157:H7 strain, EDL933, caused the first recognized EHEC outbreak in 1982 (6), and it's genome shares a common 4.1 Mb DNA backbone with the non-pathogenic laboratory strain of *E. coli* K-12, MG1655 (7–9). However, the EDL933 genome also contains 1.34Mb of chromosomal DNA that is absent from K-12, as well as a 90kb virulence plasmid pO157. EDL933-specific 'O-islands' encode genes recognized as the major EHEC virulence factors and are thought to have been acquired by horizontal gene transfer; many are encoded within putative prophage elements.

Although there are a variety of EHEC serotypes and the O-island complement in different EHEC isolates can differ (10), EHEC genomes all contain one or more prophages encoding Shiga toxins and the Locus of Enterocyte Effacement (LEE) pathogenicity island (11). These two horizontally acquired elements are critical EHEC virulence determinants. Shiga toxins contribute to diarrhea and the development of HUS (4,5,12). The LEE encodes a type III secretion system (T3SS) and several secreted effectors. EHEC's T3SS mediates attachment of the pathogen to colonic enterocytes,

effacement of the brush border microvilli, and the formation of actin-rich pedestal-like structures underneath attached bacteria (reviewed in (13)). Once translocated into the host cell, T3SS effectors, which are encoded both inside and outside the LEE, target diverse signaling pathways and cellular processes (14,15). A functional LEE T3SS is required for EHEC intestinal colonization in animal models as well as in humans (12,13,16–20).

In addition to the virulence factors that prompt the key symptoms of infection, EHEC also relies on bacterial factors that enable pathogen survival in and adaptation to the host environment. During colonization of the human GI tract, EHEC encounters multiple host barriers to infection, including, but not limited to stomach acid, bile, and other host- and microbiota-derived compounds with antimicrobial properties (reviewed in (21)). EHEC is known to detect intestinal cues derived from the host and the microbiota to activate expression of virulence genes and to modulate gene expression both temporally and spatially (22–25). However, comprehensive analyses of bacterial factors that contribute to EHEC survival within the host have not been reported.

The development of transposon-insertion sequencing (TIS, aka TnSeq, InSeq, TraDis, or HITS) (26–29) facilitated high-throughput and genome-scale analyses of the genetic requirements for bacterial growth in different conditions, including in animal models of infection (30–39). In this approach, the relative abundance of transposon-insertion mutants within high-density transposon-insertion libraries provides insight into loci's contributions to bacterial fitness in different environments (40,41). Potential insertion sites

for which corresponding insertion mutants are not recovered frequently correspond to regions of the genome that are required for bacterial growth (often termed “essential genes”), although the absence of a particular insertion mutant does not always reflect a critical role for the targeted locus in maintaining bacterial growth (42,43). Comparative analyses of the abundance of mutants in an initial (input) library and after growth in a selective environment (e.g., an animal host) can be used to gauge loci’s contributions to fitness in the selective condition.

Here, high-density transposon libraries were created in EHEC EDL933 and the commensal *E. coli* K-12 and used to characterize their respective in vitro growth requirements. The EHEC library was also passaged through an infant rabbit model to identify genes required for intestinal colonization. Our data indicate that during infection of the gastrointestinal tract, EHEC populations undergo a severe infection bottleneck that complicates identification of genes with true in vivo fitness defects. We used two complementary analytic approaches to circumvent the noise introduced by restrictive bottlenecks to identify genes required for colonization of the colon. More than 240 genes were found to contribute to efficient colonization of the rabbit colon. As expected, these included the LEE-encoded T3SS and *tir*, a LEE-encoded effector necessary for intestinal colonization (44). In addition, 2 non-LEE effectors and many additional new genes that encode components of the bacterium’s metabolic pathways and stress response systems were found to enable bacterial colonization of the colon. Isogenic mutants for 17 loci, including *cvpA*, a gene necessary for intestinal colonization by diverse enteric pathogens (40,45,46), were constructed, validated in the infant rabbit model and tested in vitro under

136 stress conditions that model host-derived challenges encountered within the
137 gastrointestinal tract. *cvpA* was found to be specifically required for resistance to the bile
138 salt deoxycholate and therefore appears to be a previously unappreciated member of the
139 bile-resistance repertoire of diverse enteric pathogens.
140
141 **Results and Discussion**
142
143 **Identification of genes required for EHEC growth in vitro**
144 The mariner-based Himar1 transposon, which inserts specifically at TA dinucleotides
145 (reviewed in (47)) was used to generate a high-density transposon-insertion library in a
146 Δ*lacI*::*aacC1* (gentamicin-resistance cassette inserted at *lacI*) derivative of EDL933. The
147 library was characterized via high-throughput sequencing of genomic DNA flanking sites
148 of transposon-insertion. To map the reads, we used the most recent EDL933 genome
149 sequence (8) and annotation (NCBI, February 2017). Since this genome, unlike the
150 initial EHEC genome (7), has not been linked to functional information (e.g., the EHEC
151 KEGG database, KEGG reviewed in (48)), we generated a correspondence table in
152 which the new genome annotations (RS locus tags) are linked to the original “Z
153 numbers” (Table S1). This correspondence table enabled us to utilize historically
154 valuable resources as well as the updated genomic sequence and should also benefit
155 the EHEC research community. 137,805 distinct insertion mutants were identified,
156 which corresponds to 52.5% of potential insertion mutants with an average of ~21 reads
157 per genotype (Fig S1A). Sensitivity analysis revealed that nearly all mutants were
158 represented within randomly selected read pools containing ~2 million reads. Increasing

159 sequencing depth to ~3 million reads had a negligible effect on library complexity,
160 suggesting that a sequencing depth of ~3 million reads is sufficient to identify virtually all
161 constituent genotypes within this EHEC library (Fig S1A).
162
163 EHEC's 6032 annotated genes were binned according to the percentage of disrupted
164 TA sites within each gene, and the number of genes corresponding to each bin was
165 plotted (Fig 1A). As expected for a high-density transposon-insertion library (41), this
166 distribution was bimodal, with a minor peak comprised of genes disrupted in few
167 potential insertion sites (Fig 1A, left), and a major peak comprised of genes that are
168 disrupted in most or all potential insertion sites (Fig 1A, right). Based on the center of
169 the major right-side peak, we estimate that ~70% of non-essential insertion sites have
170 been disrupted in this EHEC library, a degree of complexity that enabled high-resolution
171 analysis of transposon-insertion frequency.
172
173 Further analysis of insertion site distribution was performed using a hidden Markov
174 model-based analysis pipeline (EL-ARTIST, see methods and (40) for detail), that
175 classifies loci with a low frequency of transposon-insertion across the entire coding
176 sequence as 'underrepresented' (often referred to as 'essential' genes) or across a
177 portion of the coding sequence as 'regional' (Fig 1B). All other loci are classified as
178 'neutral'. Of EHEC's 6032 genes, 895 genes were classified as underrepresented, 407
179 as regional, and 4,730 as neutral (Fig 1A, Table S2). In general, neutral genes (blue)
180 were disrupted in a higher percentage of TA sites than underrepresented genes (red) or
181 regional genes (purple), which displayed low and intermediate percentage of TA site

182 disruption, respectively (Fig 1B). Neutral genes are considered to be dispensable for
183 growth in LB, whereas non-neutral genes (regional and underrepresented genes
184 combined) likely have important functions for growth in this media or are otherwise
185 refractory to transposon insertion (42,43).
186
187 We identified Z Numbers (Table S1) and the linked Clusters of Orthologous Groups
188 (COG) (49,50) and KEGG pathways associated with the 1302 genes classified as non-
189 neutral (underrepresented and regional). Although some loci have no COG assigned,
190 714 genes were assigned a COG functional category (Table S2). Each COG category
191 was plotted against its “COG Enrichment Index”, which is calculated as the percentage
192 of non-neutral genes in each COG category divided by the percent of the whole genome
193 with that COG (51). A subset of COGs, particularly cell cycle control, translation, lipid
194 and coenzyme metabolism, and cell wall biogenesis were associated with non-neutral
195 genes at a frequency significantly higher than expected based on their genomic
196 representation (Fig S2A). Collectively, the COG and a similar KEGG analysis (Table S3)
197 revealed that EHEC genes with non-neutral transposon-insertion profiles are associated
198 with pathways and processes often linked to essential genes in other organisms (52).
199
200 Non-neutral genes comprise ~22% of EHEC’s annotated genes, a proportion of the
201 genome that is substantially larger than the 8% and 9% observed in analogous TIS-
202 based characterizations of *Vibrio cholerae* and *Vibrio parahaemolyticus* (40,45). Of the
203 1,302 non-neutral EHEC genes, only 760 are homologous to a gene in *E. coli* K-12
204 MG1655 (>90% nucleotide identity or >90% amino acid identity across 90+% of gene

length) (Table S2, column M); thus, EDL933-specific loci comprise a high proportion (~42%) of EHEC's underrepresented loci. The enrichment of underrepresented loci among EHEC-specific genes, many of which were acquired by horizontal gene transfer, may reflect factors that can limit transposon-insertion other than fitness costs.

Previous analyses revealed that nucleoid binding proteins such as HNS, which binds to DNA with low GC content, can hinder Himar1 insertion (42). Consistent with this observation, genes classified as non-neutral have a lower average GC content than genes classified as neutral (Fig S2B; blue vs red distributions). The disparity in GC content between neutral and non-neutral loci is particularly marked for EHEC genes that do not have a homolog in K-12 (divergent; Fig 1C), although there is also a significant difference between the GC content of neutral and non-neutral loci with a K-12 homolog (homolog; Fig 1C). These analyses suggest that there is an association between GC content and transposon-insertion frequency in EHEC, as in other organisms, and that the prevalence of underrepresented loci among divergent loci may in part stem from the lower average GC content of these loci (Fig S2C). Additional studies are necessary to determine if the association between low GC content and reduced transposon-insertion is due to HNS-binding, other nucleoid-associated proteins, or as yet unidentified fitness-independent transposon insertion biases.

TIS-based comparison of EHEC and *E. coli* in vitro growth requirements

To evaluate whether the abundance of non-neutral loci was specific to EHEC or was characteristic of additional *E. coli* strains, a high-density transposon-insertion library was

constructed in a Δ*lacI*::*cat* (chloramphenicol-resistance cassette inserted at *lacI*) derivative of *E. coli* K-12 MG1655. EL-ARTIST analysis of the high-density K-12 library (Fig S1B) was implemented with the same parameters as those for the EHEC library analysis and classified 24% of genes as underrepresented (786 underrepresented, 300 regional and 3397 neutral; Fig 1D, Table S4). Comparison of the gene classification of homologous loci (Table S2 vs Table S4) revealed substantial concordance between the sets of genes with non-neutral insertion profiles in EHEC and K-12: 83% (629/760) of the non-neutral EHEC genes with homologs in K-12 were likewise classified as non-neutral in the *E. coli* K-12 strain (Fig 1E). Thus, analyses of non-neutral loci suggest either that the majority of ancestral loci make similar contributions to the survival and/or proliferation of EHEC and K-12 strains in LB or that they are similarly resistant to transposon-insertion.

We further explored the 131 underrepresented EHEC loci (Table S5) that were classified as neutral (able to sustain insertions) in *E. coli* K-12. Most of these genes have are linked to KEGG pathways for metabolism, particularly metabolism of galactose, glycerophospholipid, and biosynthesis of secondary metabolites (Table S5, Fig 1F). While this divergence could reflect the laboratory adaptation of the K-12 isolate, gene acquisition during EHEC evolution may have heightened the pathogen's reliance on metabolic processes that are not critical for growth of K-12. Such ancestral genes may be useful targets for antimicrobial agents, as they might antagonize EHEC growth without disruption of closely related commensal Enterobacteriaceae populations.

Comparison of TIS and deletion-based gene classification

The sets of genes classified as underrepresented or regional in EHEC and K-12 transposon libraries were compared to the 300 genes classified as essential in the K-12 strain BW25113 based on their absence from a comprehensive library of single gene knockouts (53–55). 98% of these genes (294/300) were also classified as underrepresented or regional in EDL933 and MG1655 (Table S2 and S4). The few loci previously classified as essential but not found to be underrepresented or regional in our analysis include several small genes, whose low number of TA sites hampers confident classification. One gene in this list, *kdsC*, was found to have insertions across the gene in both EDL933 and MG1655 (Fig S2D). *kdsC* knockouts have also been reported previously (56), confirming that this locus is not required for K-12 growth despite the absence of an associated mutant within the Keio collection. Thus, underrepresented and regional loci encompass, but are not limited to, loci previously classified as essential.

Several factors likely account for over-estimation of loci as underrepresented or regional. First, loci can be classified as underrepresented even when viable mutants are clearly present within the insertion library (Fig 1A); insertions simply need to be consistently less abundant across a segment of the gene than insertions at other (neutral) sites. Loci may also be classified as underrepresented due to fitness-independent insertion biases, as discussed above (42,43). Additional evidence that loci categorized as non-neutral by transposon-insertion studies are not necessarily essential for growth was provided by a recent study of essential genes in E. coli K-12 (57).

274 However, the more expansive non-neutral classification can provide insight into loci that
275 enable optimal growth, in addition to those that are required.
276
277 **Identification of EHEC genes required for growth in vivo**
278 To identify mutants deficient in their capacity to colonize the mammalian intestine, the
279 EHEC transposon library was orogastrically inoculated into infant rabbits, an established
280 model host for infection studies (12,44,58,59). Transposon-insertion mutants were
281 recovered from the colon at 2 days post-infection, and the sites and abundance of
282 transposon-insertion mutations were determined via sequencing, as described above.
283 The relative abundance of individual transposon-insertion mutants in the library
284 inoculum was compared to samples independently recovered from the colons of 7
285 animals to identify insertion mutants that were consistently less abundant in libraries
286 recovered from the colon. Under ideal conditions, this signature is indicative of negative
287 selection of the mutant during infection, reflecting that the disrupted locus is necessary
288 for optimal growth within the intestine.
289
290 Sequencing and sensitivity analyses of the 7 passaged libraries revealed that they
291 contained substantially fewer unique insertion mutants than the library inoculum (23-
292 38% total mutants recovered, ~30,000 of 120,000) (Fig S1C-J). These data are
293 suggestive of population constrictions that could have arisen from 2 distinct but not
294 mutually exclusive causes: 1) negative selection, leading to depletion of mutants
295 deficient at in vivo survival or intestinal colonization; and/or 2) infection bottlenecks,
296 population constrictions that lead to stochastic reductions in the average number of

297 insertions per gene, independent of genotype or selective pressures. We binned genes
298 according to the percentage of TA sites disrupted within their gene sequences and
299 plotted the number of genes corresponding to each bin for both the inoculum (Fig 2A-
300 top) and a representative rabbit-passaged sample (Fig 2A-bottom). The passaged
301 sample exhibited a marked leftward shift relative to the inoculum, a signature indicative
302 of population constriction due to an infection bottleneck (41,60).
303
304 The TIS data was further analyzed using the Con-ARTIST pipeline. Con-ARTIST uses
305 iterative simulation-based normalization to compensate for experimental bottlenecks to
306 facilitate discrimination between stochastic reductions in genotype abundance and
307 reductions attributable to bona fide negative selection (mutants for which there was a
308 fitness cost in the host environment) (40). The Con-ARTIST analysis protocol and
309 subsequent gene classification is schematized in Figure 2B. For libraries recovered
310 from each rabbit, we used this workflow to classify genes as ‘conditionally depleted’
311 (red, CD), ‘queried’(blue) or ‘insufficient data’ (black) (Fig 2B) compared to the inoculum
312 library. CD genes contain sufficient insertions for analysis (see methods) and meet a
313 standard of a 4-fold reduction in read abundance that is consistent across TA sites in a
314 gene (Fig 2B). Queried genes contained sufficient insertions for analysis but failed to
315 meet this criterion. Genes classified as insufficient do not contain sufficient insertions for
316 analysis. The output of gene categorization using these thresholds is displayed for a
317 single rabbit in Fig 2C (additional animals in Fig S3A-G) and summarized for all animals
318 in Fig 2D.
319

The restrictive bottleneck and animal to animal variation led to differences in the numbers of genes classified as CD in different rabbits (Fig 2D). Due to this variability, an additional criterion that genes be classified as CD in 5 or more of the 7 animals analyzed was imposed to create a consensus cutoff. In contrast to the >2000 genes classified as conditionally depleted in one or more animals, only 243 genes were classified as conditionally depleted across 5 or more animals (Fig S3H, Table S6). These relatively stringent standards were imposed in order to identify robust candidates for genes that facilitate EHEC intestinal growth, despite the limitations of the infection bottleneck in this experimental model. Therefore, we do not conclude that genes classified as “Queried” (3860) or “Insufficient Data” (1926) are *not* attenuated relative to the wild type strain in vivo; it is likely that the list of CD loci (Table S6) is incomplete.

Using our Z correspondence table (Table S1), 89% (217/243) genes classified as conditionally depleted were assigned to a COG functional category. CD genes were frequently associated with amino acid and nucleotide metabolism, signal transduction, and cell wall/envelope biogenesis, but only amino acid metabolism reached statistical significance after correction for multiple hypothesis testing (Fig 2E). These genes are also associated with KEGG metabolic pathways (particularly amino acid metabolism), several two-component systems, including *qseC*, which has previously been implicated in EHEC virulence gene regulation, and lipopolysaccharide biosynthesis (22) (Table S7, Fig 2F). 33 of the 243 CD genes are EHEC specific, whereas the remaining 210 have homologs in K-12 (Table S6), highlighting the importance of conserved metabolic pathways in the pathogen’s capacity to successfully colonize its colonic niche. Similar

343 metabolic pathways were also found to be important for *V. cholerae* growth in the infant
344 rabbit small intestine (40,61), and raising the possibility of targeting metabolic pathways
345 such as those for amino acid biosynthesis with antibiotics (62–65).
346
347 The stochastic loss of individual insertion mutants in severely bottlenecked data can
348 hinder Con-ARTIST-based identification of CD genes. In particular, meeting the
349 pipeline's consistency threshold as measured by the Mann Whitney U (MWU) p-value
350 (Fig 2B) is difficult because severe bottlenecks drastically decrease the number of
351 individual transposon-insertion mutants per gene; therefore, the mutant replicates
352 needed to demonstrate consistency are not present. Queried genes often did not meet
353 the MWU p-value cut-off, even though fold change information may suggest marked
354 attenuation. To reclaim some of the mutants that were not classifiable by Con-ARTIST,
355 we also used Comparative TIS (CompTIS), a principal components analysis (PCA)-
356 based framework (66), to compare the 7 libraries recovered from rabbit colons. PCA is a
357 dimensional reduction approach used to describe the sources of variation in multivariate
358 datasets. Recently, we found that PCA is useful for identifying genes whose inactivation
359 leads to mutant growth phenotypes that are consistent across TIS replicates (66). Here,
360 we applied CompTIS as an alternative approach to identify genes with phenotypic
361 consistency (inability to colonize the rabbit colon) in all 7 rabbit replicates.
362
363 To perform CompTIS, the fold change of each gene from the seven colon libraries was
364 subjected to gene level PCA (glPCA) (see methods and (66)), with each library
365 recovered from a rabbit colon representing a replicate. glPC1 describes most of the

366 variation in the animals (Fig S3I) and reports a weighted average of the fold change
367 values for each gene across the 7 animals (Table S6). The signs and magnitudes of
368 PC1 were all similar (Fig S3J), indicating that each rabbit contributes approximately
369 equally to PC1, as expected for biological replicates. The distribution of glPC1 scores is
370 continuous (Fig 2G) and describes each gene's contribution to EHEC intestinal
371 colonization. Most genes have a glPC1 score close to zero (average PC1=0),
372 suggesting that they do not contribute to colonization. However, the distribution includes
373 a left tail beginning at PC1 scores of approximately -900 that encompassed the lowest
374 10% of scores (Fig 2G, Fig S3K), which likely correspond to genes contributing to
375 colonization. This list of 541 genes included nearly all (85%) of the genes classified as
376 CD by the more conservative Con-ARTIST analysis outlined above (Fig 2B). This
377 method allows for identification of additional candidate genes required for
378 survival/growth in vivo. For example, the PCA approach captured genes such as *ler*
379 (PC1 = -2290), a critical activator of the LEE T3SS (67), which was classified as queried
380 by Con-ARTIST due to the relative paucity of unique insertion mutants.
381
382 **Analyses of the requirement for T3SS and its associated effectors in colonization**
383 To begin to assess the accuracy of our gene classifications using the Con-ARTIST
384 consensus approach and CompTIS, we examined classifications within the LEE
385 pathogenicity island, which encodes the EHEC T3SS and plays a critical role in
386 intestinal colonization (12,16–19). The LEE is comprised of 40 genes, including genes
387 encoding the structural components of the T3SS, some of the pathogen's effectors, their
388 chaperones, and Intimin (encoded by *eae*), the adhesin that binds to the translocated

389 Tir protein. In infant rabbits, previous studies using single deletion mutants revealed that
390 *tir, eae*, and *escN*, the T3SS ATPase, were all required for colonization (12,44). We
391 observed a marked reduction in the abundance of insertions across nearly the entire
392 LEE in the samples from the rabbit colons relative to the simulation-normalized input
393 reads (Fig 3A). The 3 genes previously found to be required for colonization (*tir*, *eae*
394 and *escN*) were classified as CD using the Con-ARTIST consensus approach,
395 enhancing confidence in this scheme. Furthermore, 8 additional LEE-encoded genes
396 critical for T3SS activity, including translocon T3SS components (*espB*, *espD*, and
397 *espA*) and structural components (*escD*, *escQ*, *escV*, *escI*, and *escC*) were also
398 classified as CD using this scheme (Fig 3AB, Table S6) (16,68–72). Our findings
399 provide additional strong evidence that the LEE T3SS is critical for EHEC proliferation in
400 the intestine. However, many LEE-encoded genes had insufficient data to enable
401 classification via the Con-ARTIST consensus approach.
402
403 In contrast to Con-ARTIST analytic approach, with the PCA-based CompTIS analysis,
404 we were able to assess the contribution of all of the genes in the LEE and identify
405 several more genes likely to be important for in vivo colonization. With this approach,
406 most LEE genes had glPC1 scores in the bottom 10% of the distribution (Table S6, Fig
407 3B). Notably, the genes that were not in this portion of the distribution included 4
408 effectors (*espF*, *map*, *espH*, *espG*). Previous studies in infant rabbits showed that *espG*
409 and *map* were dispensable for colonic colonization and that *espH* and *espF* mutants
410 only had modest colonization defects (44), lending credence to PCA-based
411 classification.

412

413 EHEC has a large suite of non-LEE encoded effectors (Nle), many of which reside
414 within prophage elements. Only 2 of 43 Nle genes (*nleA, espM1)* were classified as CD
415 by Con-ARTIST or were found within the bottom 10% of glPC1 scores by CompTIS
416 (Fig3B, Table S6), suggesting that only a small subset of EHEC effectors are critical for
417 colonization, while other effectors likely play auxiliary roles. NleA was previously
418 reported to be important for colonic colonization by a related enteric pathogen,
419 *Citrobacter rodentium* (73), and is thought to suppress inflammasome activity (74), while
420 EspM1 is thought to modulate host actin cytoskeletal dynamics (75,76). Additional
421 studies are warranted to confirm and further explore how these 2 effectors play pivotal
422 roles promoting intestinal colonization.

423

424 **Validation of colonization defects in non-LEE encoded genes classified as**
425 **conditionally depleted**

426 We performed further studies of 17 conditionally depleted genes/operons that had not
427 previously been demonstrated to promote EHEC intestinal colonization. The Con-
428 ARTIST consensus approach and CompTIS classified all of these genes as
429 conditionally depleted except one, *hupB*, which was classified as queried by Con-
430 ARTIST but within the bottom 10% of glPC1 scores (Fig 4A). Mutants with in-frame
431 deletions of either single loci (*agaR*, *cvpA*, *envC, htrA*, *hupB, mgtA*, *oxyR*, *prc*, *sspA*,
432 *sufI*, *tolC*, and RS09610, a hypothetical gene of unknown function) or operons with one
433 or more genes classified as conditionally depleted (*acrAB*, *clpPX*, *envZompR*, *phoPQ*,
434 *tatABC*) were generated. Then, each mutant strain was barcoded with unique sequence

435 tags integrated into a neutral locus in order to enable multiplexed analysis. The in vitro
436 growth of the barcoded mutants was indistinguishable from that of the WT strain (Fig
437 S4A), suggesting that the transposon mutants' in vivo attenuation is not explained by a
438 generalized growth deficiency.
439
440 The barcoded mutants, along with the barcoded WT EHEC, were co-inoculated into
441 infant rabbits to compare the colonization properties of the mutants and WT. The
442 relative frequencies of WT and mutant EHEC within CFU recovered from infected
443 animals was enumerated by deep sequencing of barcodes, and these frequencies were
444 used to calculate competitive indices (CI) for each mutant (i.e., relative abundance of
445 mutant/WT tags in output normalized to input). 14 of the 17 mutants tested had CI
446 values significantly lower than 1, validating the colonization defects inferred from the
447 TIS data (Fig 4). In aggregate, these observations support our experimental and
448 analytical approaches and suggest that many of the genes classified as CD by the Con-
449 ARTIST consensus approach and/or have low PC-1 scores may also contribute to
450 intestinal colonization.
451
452 **Many conditionally depleted loci exhibit reduced T3SS effector translocation**
453 **and/or increased sensitivity to extracellular stressors**
454 The many new genes implicated in EHEC colonization by the TIS data could contribute
455 to the pathogen's survival and growth in vivo by a large variety of mechanisms. Given
456 the pivotal role of EHEC's T3SS in intestinal colonization, as well as previous
457 observations that factors outside the LEE can regulate T3SS gene expression and/or

458 activity (reviewed in (67)), we assessed whether T3SS function was impaired in the 11
459 mutants with CIs <0.3 (Fig 4). Translocation of EspF (an effector protein) fused to a
460 TEM-1 beta-lactamase reporter into HeLa cells was used as an indicator of T3SS
461 functionality (77). An Δ*escN* mutant, which lacks the ATPase required for T3SS
462 function, was used as a negative control.
463
464 Deletions in three protease-encoded genes, *clpPX*, *htrA*, and *prc*, were associated with
465 reduced EspF translocation (Fig 5A). Both ClpXP and HtrA have been implicated in
466 T3SS expression/activity in previous reports (78–81). The ClpXP protease controls LEE
467 gene expression indirectly by degrading LEE-regulating proteins RpoS and GrlR (82).
468 The periplasmic protease HtrA (aka DegP) has been implicated in post-translational
469 regulation of T3SS as part of the Cpx-envelope stress response (80,81). Interestingly,
470 *prc,* which also encodes a periplasmic protease (82), also appears required for robust
471 EspF translocation. Prc has been implicated in the maintenance of cell envelope
472 integrity under low and high salt conditions in *E. coli* K-12 (83). Consistent with this
473 observation, in high osmolarity media a Δ*prc* EHEC mutant exhibited cell shape defects
474 (Fig S4B). Deficiencies in the cell envelope associated with absence of Prc may impair
475 T3SS assembly and/or function, perhaps also by triggering the Cpx-envelope stress
476 response. Together, these observations suggest that in vivo these three proteases
477 modulate T3SS expression/function, thereby promoting EHEC intestinal colonization.
478
479 We also investigated the capacity of each of the 11 mutant strains to survive challenge
480 with three stressors – low pH, bile, and high salt (osmotic challenge) – that the

481 pathogen may encounter in the gastrointestinal tract. Relative to the WT strain, all but
482 one (*sufI*) of the mutant strains exhibited reduced survival following one or more of
483 these challenges (Fig 5BC), suggesting that exposure to these host environmental
484 factors may contribute to the in vivo attenuation of these mutants. Many of the EHEC
485 mutants exhibited sensitivities to external stressors that are consistent with previously
486 described phenotypes in other organisms and experimental systems. For example, the
487 EHEC Δ*acrAB* locus, which was associated with bile sensitivity in EHEC (Fig 5B), is
488 known to contribute to a multidrug efflux system that can extrude bile salts, antibiotics,
489 and detergents (84). Our observation that mutants lacking the oxidative stress response
490 gene *oxyR* are sensitive to bile and to acid pH is also concordant with previous reports
491 linking both stimuli to oxidative stress (85–87). Furthermore, the heightened sensitivity
492 to bile, acid, and elevated osmolarity of EHEC lacking the two-component regulatory
493 system EnvZ/OmpR is consistent with previous reports that EnvZ/OmpR is a critical
494 determinant of membrane permeability, due to its regulation of outer membrane porins
495 OmpF and OmpC. Mutations that activate this signaling system (in contrast to the
496 deletions tested here) have been found to promote *E. coli* viability in vivo and to
497 enhance resistance to bile salts (88).
498
499 The EHEC Δ*tatABC* mutant exhibited a marked colonization defect and a modest
500 increase in bile sensitivity. The twin-arginine translocation (Tat) protein secretion
501 system, which transports folded protein substrates across the cytoplasmic membrane
502 (reviewed in (89,90)), has been implicated in the pathogenicity of a variety of Gram-
503 negative pathogens, including enteric pathogens such as *Salmonella enterica* serovar

504 Typhimurium (91–93), *Yersinia pseudotuberculosis* (94,95), *Campylobacter jejuni* (96),
505 and *Vibrio cholerae* (97). Attenuation of Tat mutants can reflect the combined absence
506 of a variety of secreted factors. For example, the virulence defect of *S. enterica*
507 Typhimurium *tat* mutants are likely due to cell envelope defects caused by the inability
508 to secrete the periplasmic cell division proteins AmiA, AmiC and SufI (92). Notably,
509 single knock-outs of any of these genes did not cause attenuation (92), but altogether
510 their absence renders the cell-envelope defective and more sensitive to cell-envelope
511 stressors, such as bile acids (93).
512
513 In EHEC, the Tat system has been implicated in Stx1 export (98), but since Stx1 was
514 not a hit in our screen and is not thought to modulate intestinal colonization (12), it is not
515 likely to explain the marked colonization defect of the EHEC Δ*tatABC* mutant. The suite
516 of EHEC Tat substrates has not been experimentally defined, although putative Tat
517 substrates can be identified by a characteristic signal sequence (89,90). A few
518 substrates, including SufI, OsmY, OppA, MglB, and H7 flagellin, have been detected
519 experimentally (98). *sufI,* interestingly, was also a validated hit in our screen and is the
520 only Con-ARTIST defined CD gene that has a predicted Tat-secretion signal. However,
521 the Δ*sufI* mutant did not display enhanced bile sensitivity, suggesting that attenuation of
522 this mutant, and perhaps of the Δ*tatABC* mutant as well, reflects deficiencies in other
523 processes. SufI is a periplasmic cell division protein that localizes to the divisome and
524 may be important for maintaining divisome assembly during stress conditions (99,100).
525 *E. coli tat* mutants have septation defects (101), presumably from loss of SufI at the
526 divisome. Interestingly, *envC*, another validated CD gene, encodes a septal murein

527 hydrolase (102) that is required for cell division, and the Δ*envC* mutant also displayed
528 increased bile sensitivity. Consistent with this hypothesis, in high osmolarity media, the
529 Δ*sufI,* Δ*envC,* and Δ*tatABC* mutants exhibited septation or cell shape defects (Fig S4B).
530 Collectively, these data suggest that an impaired capacity for cell division may reduce
531 EHEC's fitness for intraintestinal growth, and that at times this may reflect increased
532 susceptibility to clearance by host factors such as bile.
533
534 **CvpA promotes EHEC resistance to deoxycholate**
535 We further characterized EHEC Δ*cvpA* because other TIS-based studies of the
536 requirements for colonization by diverse enteric pathogens (*Vibrio cholerae*, *Vibrio*
537 *parahaemolyticus* and *Salmonella enterica* serovar Typhimurium) also classified *cvpA*
538 as important for colonization, but did not explore the reasons for the colonization
539 deficiency of the respective mutants (40,45,46).
540
541 *cvpA* encodes a putative inner membrane protein and has been linked to colicin V
542 export in *E. coli* K-12 (103) as well as curli production and biofilm formation in UPEC
543 (104). The EHEC Δ*cvpA* mutant did not exhibit an obvious defect in biofilm formation or
544 curli production (Fig S5AB), suggesting that *cvpA* may have a distinct role in EHEC
545 pathogenicity.
546
547 To further characterize the sensitivity of EHEC Δ*cvpA* mutant to bile, we exposed the
548 mutant to the two major bile salts found in the gastrointestinal tract, cholate (CHO) and
549 deoxycholate (DOC) (Fig 6AC) (85,105). Cholate is a primary bile salt, produced in the

liver and released into the biliary tract, while deoxycholate is a secondary bile salt that is generated from cholate by intestinal bacteria in the colon. In contrast to WT EHEC, which displayed equivalent sensitivity to the two bile salts in MIC assays (MIC = 2.5% for both), the Δ*cvpA* mutant was much more sensitive to DOC than to CHO (MIC= 0.08% versus 1.25%). The Δ*cvpA* mutant's sensitivity to deoxycholate was present both in liquid cultures and during growth on solid media (Fig. 6ABC).

Growth of the Δ*cvpA* mutant in the presence of deoxycholate was partially restored by introduction of *cvpA* under the control of an inducible promoter, confirming that sensitivity is linked to the absence of *cvpA* (Fig. 6A). *cvpA* lies upstream of the purine biosynthesis locus *purF,* and some Δ*cvpA* mutant phenotypes have been attributed to reduced expression on *purF* due to polar effects (103,106). The growth of the EHEC Δ*cvpA* mutant was not impaired in the absence of exogenous purines (Fig S5C), suggesting the *cvpA* deletion does not adversely modify *purF* expression.

Bile sensitivity has been associated with defects in the bacterial envelope or with reduced efflux capacity (reviewed in (105)). We assessed the growth of the Δ*cvpA* mutant in the presence of a variety of agents that perturb the cell envelope to assess the range of the defects associated with the absence of *cvpA*. The MICs of WT and Δ*cvpA* EHEC were compared to those of an Δ*acrAB* mutant, whose lack of a broad-spectrum efflux system provided a positive control for these assays. Notably, the Δ*cvpA* mutant did not exhibit enhanced sensitivity to any of the compounds tested other than bile salts. In marked contrast, the Δ*acrAB* mutant displayed increased sensitivity to all

573 agents assayed (Fig 6C). These observations suggest that the sensitivity of the *cvpA*
574 mutant to DOC is not likely attributable to a general cell envelope defect in this strain. *V.*
575 *cholerae* and *V. parahaemolyticus* Δ*cvpA* mutants also exhibited sensitivity to DOC (Fig
576 S5D), implying a similar role in bile resistance in these distantly related enteric
577 pathogens.
578
579 A variety of bioinformatic algorithms (PSLPred, HHPred, Phobius, Phyre2) suggest that
580 CvpA is an inner membrane protein with 4-5 transmembrane elements similar to small
581 solute transporter proteins (Fig 6D). Phyre2 and HHPred reveal CvpA's partial similarity
582 to inner membrane transporters in the Major Facilitator Superfamily of transporters
583 (MFS) and the small-conductance mechanosensitive channels family (MscC). Additional
584 protein classification schemes group CvpA with proteins involved in solute transport. For
585 example, the PFAM database groups CvpA (PF02674) in the LysE transporter
586 superfamily (CL0292), a set of proteins known to enable solute export. In conjunction
587 with findings presented above, these predictions raise the possibility that CvpA is
588 important for the export of a limited set of substrates that includes DOC. Additional
589 studies to confirm this hypothesis and to establish how CvpA enables export are
590 warranted, particularly because this protein is widespread amongst enteric pathogens.
591
592 **Conclusions**
593 Here, we created a highly saturated transposon library in EHEC EDL933 to identify
594 the genes required for in vitro and in vivo growth of this important food-borne pathogen
595 using TIS. This approach has transformed our capacity to rapidly and fairly

596 comprehensively assess the contribution an organism's genes to growth in different
597 environments (41,107,108). However, technical and biologic issues can confound
598 interpretation of genome-scale transposon-insertion profiles. For example, we found
599 that EHEC genes with low GC content or those without homologs in K-12 were less
600 likely to contain transposon-insertions (Fig S2BC, Fig1C). Many of these genes were
601 likely acquired during EHEC evolution via lateral gene transfer; they constitute some of
602 the ~1.4MB of DNA that distinguishes EHEC from K-12 strains. Unexpectedly, more
603 than 100 of the genes conserved between EHEC and K-12 appear to promote the
604 growth of the pathogen in rich media but not that of K-12 (Table S5), suggesting that the
605 ~1.4MB of laterally acquired DNA that distinguishes EHEC and K-12 has enabled
606 divergence of the metabolic roles of ancestral *E. coli* genes in these backgrounds.
607
608 In animal models of infection, bottlenecks that result in marked stochastic loss of
609 transposon mutants can severely constrain TIS-based identification of genes required
610 for in vivo growth. Analysis of the distributions of the EHEC transposon-insertions in
611 vitro and in vivo (Fig. 2) revealed that there is a large infection bottleneck in the infant
612 rabbit model of EHEC colonization. Both Con-ARTIST, which applies conservative
613 parameters to define conditionally depleted genes (Fig 2B), and a PCA-based
614 approach, CompTIS, were used to circumvent the analytical challenges posed by the
615 severe EHEC infection bottleneck. These approaches should also be of use for similar
616 bottlenecked data that often hampers interpretation of TIS-based infection studies.
617 Validation studies, which showed that 14 of 17 genes (82%) classified as CD were
618 attenuated for colonization, suggest that these approaches are useful. Besides the LEE-

619 encoded T3SS, more than 200 additional genes were found to contribute to EHEC
620 survival and/or growth within the intestine. This set of genes should be of considerable
621 value for future studies elucidating the processes that enable the pathogen to proliferate
622 in vivo and for design of new therapeutics.
623
624 **Materials and Methods**
625
626 **Ethics statement**
627 All animal experiments were conducted in accordance with the recommendations in the
628 Guide for the Care and Use of Laboratory Animals of the National Institutes of Health
629 and the Animal Welfare Act of the United States Department of Agriculture using
630 protocols reviewed and approved by Brigham and Women's Hospital Committee on
631 Animals (Institutional Animal Care and Use Committee protocol number 2016N000334
632 and Animal Welfare Assurance of Compliance number A4752-01)
633
634 **Bacterial strains, plasmids and growth conditions**
635 Strains, plasmids and primers used in this study are listed in Supplementary Tables 8
636 and 9. Strains were cultured in LB medium or on LB agar plates at 37°C unless
637 otherwise specified. Antibiotics and supplements were used at the following
638 concentrations: 20 μg/mL chloramphenicol (Cm), 50 μg/mL kanamycin (Km), 10 μg/mL
639 gentamicin (Gent), 50 μg/mL carbenicillin (Carb), and 0.3 mM diaminopimelic acid
640 (DAP).
641

642 A gentamicin-resistant mutant of *E. coli* O157:H7 EDL933 (Δ*lacI::aacC1*) and a
643 chloramphenicol-resistant mutant of *E. coli* K-12 MG1655 (Δ*lacI::cat*) were used in this
644 study for all experiments, and all mutations were constructed in these strain
645 backgrounds except where specified otherwise. The Δ*lacI::aacC1* and Δ*lacI*::*cat*
646 mutations were constructed by standard allelic exchange techniques (109) using a
647 derivative of the suicide vector pCVD442 harboring a gentamicin resistance cassette
648 amplified from strain TP997 (Addgene strain #13055) (110) or a chloramphenicol
649 resistance cassette from plasmid pKD3 (Addgene plasmid #45604) (53) flanked by the
650 5' and 3' DNA regions of the *lacI* gene. Isogenic mutants of EDL933 Δ*lacI::aacC1* were
651 also constructed by standard allelic exchange using derivatives of suicide vector pDM4
652 harboring DNA regions flanking the gene(s) targeted for deletion. *E. coli* MFDλpir (111)
653 was used as the donor strain to deliver allelic exchange vectors into recipient strains by
654 conjugation. Sequencing was used to confirm mutations.
655
656 A Δ*cvpA* strain was also constructed using standard allelic exchange in a streptomycin-
657 resistant mutant (Sm^R) of *V. parahaemolyticus* RIMD 2210633. A *cvpA*::tn mutant was
658 used from a *Vibrio cholerae* C6706 arrayed transposon library (112).
659
660 **Transposon-insertion library construction**
661 To create transposon-insertion mutant libraries in EHEC EDL933 Δ*lacI::aacC1*,
662 conjugation was performed to transfer the transposon-containing suicide vector pSC189
663 (113) from a donor strain (*E. coli* MFDλpir) into the EDL933 recipient. Briefly, 100 µL of
664 overnight cultures of donor and recipient were pelleted, washed with LB, and combined

665 in 20 µL of LB. These conjugation mixtures were spotted onto a 0.45 µm HA filter
666 (Millipore) on an LB agar plate and incubated at 37°C for 1 h. The filters were washed in
667 8 mL of LB and immediately spread across three 245x245 mm^2 (Corning) LB-agar
668 plates containing Gent and Kn. Plates were incubated at 37°C for 16 h and then
669 individually scraped to collect colonies. Colonies were resuspended in LB and stored in
670 20% glycerol (v/v) at -80°C as three separate library stocks. The three libraries were
671 pooled to perform essential genes analysis, and one library aliquot was used to as an
672 inoculum for infant rabbit infection studies.
673
674 To create TIS mutant libraries in *E. coli* K-12 MG1655 Δ*lacI*::cat, conjugation was
675 performed as above. 200 uL of overnight culture of the donor strain (*E. coli* MFDλpir
676 carrying pSC189) and the recipient strain (MG1655 Δ*lacI*::cat) were pelleted, washed,
677 combined and spotted on 0.45 µm HA filters at 37°C for 5.5 hours. Cells were collected
678 from the filter, washed, plated on selective media (LB Kan, Cat), and incubated
679 overnight at 30°C. Colonies were resuspended in LB and frozen in 20% glycerol (v/v).
680 An aliquot was thawed and gDNA isolated for analysis.
681
682 **Infant rabbit infection with EHEC transposon-insertion library**
683 Mixed gender litters of 2-day-old New Zealand White infant rabbits were co-housed with
684 a lactating mother (Charles River). To prepare the EHEC transposon-insertion library for
685 infection of infant rabbits, 1 mL from one library aliquot was thawed and added to 20 mL
686 of LB. After growing the culture for 3 h at 37°C with shaking, the OD_{600} was measured
687 and 40 units of culture at OD_{600}=1 (about 8 mL) were pelleted and resuspended in 10

688 mL PBS. Dilutions of the inoculum were plated on LB agar plates with Gent and Km for
689 precise dose determination. An aliquot of the inoculum was saved for subsequent gDNA
690 extraction and sequencing (input). Each infant rabbit was infected orogastrically with
691 500 µl of the inoculum ($1x10^9$ cfu) using a size 4 French catheter. Following inoculation,
692 the infant rabbits were monitored at least 2x/day for signs of illness and euthanized 2
693 days postinfection. The entire intestinal tract was removed from euthanatized rabbits,
694 and sections of the mid-colon were removed and homogenized in 1 mL of sterile PBS
695 using a minibeadbeater-16 (BioSpec Products, Inc.). 200 uL of tissue homogenate from
696 the colon were plated on LB agar + Gm + Km to recover viable transposon-insertion
697 mutants. Plates were grown for 16 h at 37°C. The next day, colonies were scraped and
698 resuspended in PBS. A 5 mL aliquot of cells was used for genomic DNA extraction and
699 subsequent sequencing (Rabbits 1-7).
700
701 **Characterization of transposon-insertion libraries**
702 Transposon-insertion libraries were characterized as described previously. Briefly, for
703 each library, gDNA was isolated using the Wizard Genomic DNA extraction kit
704 (Promega). gDNA was then fragmented to 400-600 bp by sonication (Covaris E220)
705 and end repaired (Quick Blunting Kit, NEB). Transposon junctions were amplified from
706 gDNA by PCR. PCR products were gel purified to isolate 200-500bp fragments. To
707 estimate input and ensure equal multiplexing in downstream sequencing, purified PCR
708 products were subjected to qPCR using primers against the Illumina P5 and P7
709 hybridization sequence. Equimolar DNA fragments for each library were combined and
710 sequenced with a MiSeq.

Reads were first trimmed of transposon and adaptor sequences using CLC Genomics Workbench (QIAGEN) and then mapped to *Escherichia coli* O157:H7 strain EDL933 (NCBI Accession Numbers: chromosome, NZ_CP008957.1; pO157 plasmid, NZ_CP008958.1) using Bowtie without allowing mismatches. Reads were discarded if they did not align to any TA sites, and reads that mapped to multiple TA sites were randomly distributed between the multiple sites. After mapping, sensitivity analysis was performed on each library to ensure adequate sequencing depth by sub-sampling reads and assessing how many unique transposon mutants were detected (Fig S2). Next, the data was normalized for chromosomal replication biases and differences in sequencing depth using a LOESS correction of 100,000-bp and 10,000-bp windows for the chromosome and plasmid, respectively. The number of reads at each TA site was tallied and binned by gene and the percentage of disrupted TA sites was calculated. Genes were binned by percentage of TA sites disrupted (Fig 1A, 1C).

For essential gene analysis, EL-ARTIST was used as in (45). Protein-coding genes, RNA-coding genes, and pseudogenes were included in this analysis. Briefly, EL-ARTIST classifies genes into one of three categories (underrepresented, regional, or neutral), based on their transposon-insertion profile. Classifications are obtained using a hidden Markov model (HMM) analysis following sliding window (SW) training ($p < 0.05$, 10 TA sites). Insertion-profiles for example genes were visualized with Artemis.

For identification of mutants conditionally depleted in the rabbit colon as compared to the input inoculum, Con-ARTIST was used as in (114). First, the input library was normalized to simulate the severity of the bottleneck as observed in the libraries recovered from rabbit colons using multinomial distribution-based random sampling (n=100). Next, a modified version of the Mann-Whitney U (MWU) function was applied to compare these 100 simulated control data sets to the libraries recovered from the rabbit colon. All genes were analyzed, but classification as “conditionally depleted” was restricted to genes that had sufficient data (≥5 informative TA sites), met our standard of attenuation (mean $\log_2$ fold change ≤ -2), met our standard of phenotypic consistency (MWU p-value of ≤0.05), and had a consensus classification in 5 or more of the 7 animals analyzed. Genes with ≥5 informative TA sites that fail to exceed both standards of attenuation and consistency are classified as “queried” (blue), whereas genes with less than 5 informative TA sites are classified as “insufficient data”.

Gene-level PCA (glPCA) was performed using CompTIS, a principal component analysis-based TIS pipeline, as described in (66). Briefly, $\log_2$ fold change values were derived by comparing read abundance in each sample to 100 control-simulated datasets as in Con-ARTIST. These fold change values were weighted to minimize noise due to variability (for details, see (66)). Next, genes that did not have a fold change reported for all 7 animals were discarded. The fold change values were then z-score normalized. Weighted PCA was performed in Matlab (Mathworks) with the PCA algorithm (pca).

GC content

The GC content of classified genes was compared using a Mann-Whitney U statistical test and a Bonferroni correction for multiple hypothesis correction when more than one comparison was made. A p-value <0.05 was considered significant for one comparison, p<0.025 for two. A Fisher's exact two-tailed t-test was used to compare ratios of classifications between groups, where a p-value of <0.01 was considered significant.

In vivo competitive infection

Barcodes were introduced into Δ*lacI::aacC1*and isogenic mutant strains as described previously (45,115) (46,63) Briefly, a 991bp fragment of *cynX* (RS02015) that included 51bp of the intergenic region between *cynX* and *lacA* (RS02020) was amplified using primers that contained a 30 bp stretch of random sequence and cloned into SacI and XbaI digested pGP704. The resulting pSoA176.mix was transformed into *E. coli* MFDλpir. Individual colonies carrying unique tag sequences were isolated and used as donors to deliver pSoA176 barcoded derivatives to EDL933 Δ*lacI::aacC1*and each isogenic mutant strain. Three barcodes were independently integrated into EDL933 Δ*lacI::aacC1*, and three barcodes into each isogenic mutant via homologous recombination in the intergenic region between *cynX* and *lacA*, which tolerates transposon-insertion in vitro and in vivo, indicating this locus is neutral for the fitness of the bacteria. Correct insertion of barcodes was confirmed by PCR and sequencing.

To prepare the culture of mixed EHEC-barcoded strains for the multi-coinfection experiment, 100 µl of overnight cultures of the barcoded strains were mixed in a flask

779 and 1 mL of this mix was added to 20 mL LB. After growing the culture for 3 h at 37°C
780 with shaking, the OD_{600} was measured and 40 units of culture at OD_{600}=1 (about 8 mL)
781 were pelleted and resuspended in 10 mL PBS. Dilutions of the inoculum were plated in
782 LB agar plates with Gent and Carb for precise cfu determination. 10 infant rabbits were
783 inoculated and monitored as described above, and colon samples collected. Tissue
784 homogenate was plated, and CFU were collected the following day. gDNA was
785 extracted and prepared for sequencing as in (115).
786
787 The quantification of sequence tags was done as described by (115). In brief, sequence
788 tags were amplified from the inoculum culture and libraries recovered from rabbit
789 colons. The relative in vivo fitness of each mutant was assessed by calculating the
790 competitive index (CI) as follows.
791
792 We compare two strains (Δ*lacI::aacC1* and isogenic mutant) in a population with
793 frequencies f_{wt} and $f_{mut,x}$ respectively where x is one of 17 mutant strains with a deletion
794 in gene x. For simplicity, we assume here that both expand exponentially from a time
795 point t_0 to a sampling time point t_s, their relative fitness (offspring/generation) is
796 proportional to the competitive index CI: $ln\left(\frac{f_{mut,x,s}/f_{mut,x,0}}{f_{wt,s}/f_{wt,0}}\right) = ln\,(CI)$. Here, $f_{wt,0}$ and
797 $f_{mut,x,0}$ are the frequencies of the strains in the inoculum, measured in triplicates, and $f_{wt,s}$
798 and $f_{mut,x,s}$ describe the frequencies at the sampling time point in the animal host.
799 Because the WT strain was tagged with 3 individual tags and the inoculum was

measured in triplicate, we have 3x3=9 measurements of the ratio $f_{wt,s}/f_{wt,0}$. The same is true for all mutant strains, such that we have 9 measurements of the ratio $f_{mut,x,s}/f_{mut,x,0}$. In total, we therefore have 3x3x3x3= 81 CI measurements for each mutant per animal. To determine intra-host variance in these 81 measurements, a 95% confidence interval of the CI in single animal hosts was determined by bootstrapping. For combining the CIs measured across all 10 animal hosts, we performed a random-effects meta-analysis using the metafor package (116) in the statistical software package R (version 3.0.2). The pooled rate proportions and 95% confidence intervals were calculated using the estimates and the variance of CIs in each animal determined by bootstrapping and corrected for multiple testing using the Benjamini-Hochberg procedure.

In vitro growth

Each bacterial strain was grown at 37°C overnight. The next day, cultures were diluted 1:1000 into 100 uL of LB in 96-well growth curve plates in triplicate. Plates were left shaking at 37°C for 10-24 hours. Absorbance readings at 600nm were normalized to a blank, and the average of each triplicate was taken as the optical density.

T3SS translocation assays

T3SS functionality was assessed by translocation of the known EHEC T3SS effector protein EspF into HeLa cells as described previously (77). Briefly, the plasmid encoding the effector protein EspF fused to TEM-1 beta-lactamase was transformed into each of the bacterial strains to be tested. Overnight cultures of each bacterial strain were diluted

1/50 in DMEM supplemented with HEPES (25mM), 10% FBS and L-glutamine (2mM) and incubated statically at 37°C with 5% CO2 for two hours. This media is known to induce T3SS expression (117). HeLa cells were seeded at a density of $2x10^4$ cells in 96-well clear bottom black plates and infected for 30 minutes at an MOI of 100. After 30 minutes of infection IPTG was added at a final concentration of 1mM to induce the plasmid-encoded T3SS effector. After an additional hour of incubation, monolayers were washed in HBSS solution and loaded with fluorescent substrate CCF2/AM solution (Invitrogen) as recommended by the manufacturer. After 90 minutes, fluorescence was quantified in a plate fluorescence reader with excitation at 410nm and emission was detected at 450nm. Translocation was expressed as the emission ratio at 450/520nm to normalize beta-lactamase activity to cell loading and the number of cells presented at each well, and then normalized to WT levels of translocation.

Biofilm, curli production, and purine assays

Biofilm and curli production assays were performed as described previously (104). For biofilm assays, bacterial cultures were grown in yeast extract-Casamino Acids (YESCA) medium until they reached an OD_{600} ~ 0.5 and 1/1000 dilution of this culture was used to seed 96-well PVC plates. The cultures were grown at 30°C for 48 hours and biofilm production was quantitatively measured using crystal violet staining and absorbance reading at 595nm. Relative biofilm production was normalized to the average of three WT samples. To test curli production, bacterial cultures were grown in YESCA medium until they reached an OD_{600} ~ 0.5 and then were struck to single colonies onto YESCA agar plates supplemented with Congo Red. Red colonies indicate curli production. To

846 test if our Δ*cvpA* deletion had polar effects on *purF*, the mutant and WT were struck
847 onto minimal media lacking exogenous purines.
848
849 **Acid shock assays**
850 An adaptation of the acid shock method described in (118) was performed. Briefly,
851 bacterial cultures were grown until mid-exponential phase (OD_{600} ~ 0.6), then diluted 20-
852 fold in LB pH 5.5 and incubated for 1 hour before preparing serial dilutions and plating
853 each culture to determine the relative percentage of survival in comparison to the wild-
854 type EDL933 strain. The pH of the LB broth was adjusted using sterilized 1mM HCl and
855 buffered with 10% MES. Values are expressed as percent survival normalized to WT.
856
857 **MIC assays**
858 MIC assays were performed using an adaptation of a standard methodology with
859 exponential-phase cultures (119). Briefly, the different compounds to be tested (see
860 Fig5B, 6B) were prepared in serial 2-fold dilutions in 50 ul of LB in broth in a 96-well
861 plate format. To each well was added 50 ul of a culture prepared by diluting an
862 overnight culture 1,000-fold into fresh LB broth, growing it for 1 h at 37°C, and again
863 diluting it 1,000-fold into fresh medium. The plates were then incubated without shaking
864 for 24 h at 37°C.
865
866 **Bile salts survival assays**
867 Bile salt sensitivity assays were adapted from (120). For plate sensitivity assays, each
868 bacterial strain was grown at 37°C until they reached mid-exponential phase of growth

(OD_{600nm} of 0.5) and the culture was serially diluted and spot-titered onto LB agar plates supplemented with either 1% DOC or 1% CHO. Spots were air dried and plates incubated at 37°C for 24 h. For complementation, strains were grown in media and on plates supplemented with 0.2% arabinose. For sensitivity assays done in liquid culture, each bacterial strain was grown at 37°C until it reached mid-exponential phase of growth (OD_{600nm} of 0.5) and then cultures were split and supplemented with either DOC, CHO or buffer (PBS) and bacterial growth was assessed by absorbance at 600nm.

Growth in high-salt media

Bacterial strains were grown in either LB or LB supplemented with 0.3M NaCl until mid-exponential phase and analyzed by phase microscopy.

Computational Analysis

To enable comprehensive functional/pathway analyses in EHEC we carried out BLAST-based comparisons between the old EHEC genome sequence and annotation system (NCBI Accession Numbers AE005174 and AF074613) and the new sequence and annotation system (NZ_CP008957.1 and NZ_CP008958.1) (Table S1). This comparison links the new annotations (RS locus tags) to the original 'Z numbers' from (7) and their associated function and pathway annotation.

To make the correspondence table (Table S1) between the old EHEC annotation system (Z Numbers) and the new system (RS Numbers), local BLAST was used. First, a reference nucleotide database was generated from the newest EHEC sequence and

892 annotation (NZ_CP008957.1 and NZ_CP008958.1). The EHEC genome sequence
893 containing Z number annotations (AE005174 and AF074613) was used as the query.
894 Best matches were taken as equivalent loci.
895
896 To find the K-12 homolog for EHEC genes (Table S2, column M), local BLAST was also
897 used. A reference nucleotide and amino acid database was generated from MG1655 K-
898 12 (NC_000913.3), and the newest EHEC genome sequence was used as the query.
899 For pseudogenes and genes coding for RNA, ≥90% nucleotide identity across ≥90% of
900 the gene length was considered a homolog. For protein coding genes, ≥90% amino acid
901 identity across ≥90% of the amino acid sequence was considered a homolog.
902
903 To find KEGG pathways and COG assignments for genes of interest, the Z
904 correspondence table was used to look up the Z number of each gene. The Z number
905 and corresponding functional information was searched on the EHEC KEGG database.
906
907 To determine if COGs were enriched in certain groups of genes (such as conditionally
908 depleted genes), a COG enrichment index was calculated as in (51). The COG
909 Enrichment Index is the percentage of the genes of a certain category (essential genes
910 or CD genes) assigned to a specific COG divided by the percentage of genes in that
911 COG in the entire genome. A two-tailed Fisher's exact test was used to determine if this
912 ratio was independent of grouping. A Bonferroni correction was applied for multiple
913 hypothesis testing. A p-value of <0.002 was considered to be significant.
914

Sequencing saturation of TIS libraries was determined by randomly sampling 100,000 reads from each library and identifying the number of unique mutants in that pool. Libraries are sequenced to saturation when no new mutants are identified as additional reads are added. 2-4 million reads are sufficient to capture the depth of libraries used here.

Several protein prediction programs (PSLPred, HHPred, Phobius, Phyre2) (121–123) were used to analyze the CvpA amino acid sequence. Protter (124) was used to compile information from several of these searches and generate a topological diagram. PRED-TAT (125) was used to search for tat-secretion signals in the list of CD genes.

References

1. World Health Organization. Shiga toxin-producing Escherichia coli (STEC) and food: attribution, characterization, and monitoring. Microbiological Risk Assessment Series. 2018.

2. Hartland EL, Leong JM. Enteropathogenic and enterohemorrhagic E. coli: ecology, pathogenesis, and evolution. Frontiers in Cellular and Infection Microbiology. 2015;3(15):1–3.

3. Davis KT, Van De Kar NC, Tarr PI. Shiga Toxin/Verocytotoxin-Producing Escherichia coli Infections: Practical Clinical Perspectives. Enterohemorrhagic Escherichia coli and Other Shiga Toxin-Producing E coli. 2014;321–39.

4. Karmali MA, Petric M, Steele BT, Lim C. Sporadic Cases of Haemolytic-Uraemic Syndrome Associated with Faecal Cytotoxin-producing Escherichia coli in stools. The Lancet. 1983 Mar 19;321(8325):619–20.

5. Boyce TG, Swerdlow DL, Griffin PM. Escherichia coli O157:H7 and the Hemolytic–Uremic Syndrome. New England Journal of Medicine. 1995;333(6):364–8.

6. Riley L, Remis R, Helgerson S, McGee H, Wells J, Davis B, et al. Hemorrhagic colitis associated with a rare Escherichia coli serotype. New England Journal of Medicine. 1983;308(12):681–5.

7. Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, et al. Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature. 2001;409(6819):529–33.

8. Latif H, Li HJ, Charusanti P, Palsson BO, Aziz RK. A Gapless, Unambiguous Genome Sequence of the Enterohemorrhagic Escherichia coli O157:H7 Strain EDL933. Genome Announcements. 2014;2(4).

9. Blattner F, Plunkett G, Bloch CA, Perna N, Burland V, Riley M, et al. The Complete Genome Sequence of Escherichia coli K-12. Science. 1997;277(5331):1453–62.

10. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic Escherichia coli. Proceedings of the National Academy of Sciences. 2009;106(42):17939–44.

11. Delannoy S, Beutin L, Fach P. Discrimination of Enterohemorrhagic Escherichia coli (EHEC) from Non-EHEC Strains Based on Detection of Various Combinations of Type III Effector Genes. Journal of Clinical Microbiology. 2013;51(10):3257–62.

12. Ritchie JM, Thorpe CM, Rogers AB, Waldor MK. Critical Roles for stx2, eae, and tir in Enterohemorrhagic Escherichia coli-Induced Diarrhea and Intestinal Inflammation in Infant Rabbits. Infection and Immunity. 2003;71(12):7129–39.

13. Nguyen Y, Sperandio V. Enterohemorrhagic E. coli (EHEC) pathogenesis. Frontiers in Cellular and Infection Microbiology. 2012;2(90):1–7.

14. Wong ARC, Pearson JS, Bright MD, Munera D, Robinson KS, Lee SF, et al. Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. Molecular Microbiology. 2011;80(6):1420–38.

15. Connolly JPR, Finlay BB, Roe AJ. From ingestion to colonization: the influence of the host environment on regulation of the LEE encoded type III secretion system in enterohaemorrhagic Escherichia coli. Frontiers in Microbiology. 2015;6:568.

16. Donnenberg MS, Tzipori S, McKee ML, O'Brien AD, Alroy J, Kaper JB. The role of the eae gene of enterohemorrhagic Escherichia coli in intimate attachment in vitro and in a porcine model. Journal of Clinical Investigation. 1993;92(3):1418–24.

17. Mckee ML, O'Brien AD. Investigation of Enterohemorrhagic Escherichia coli O157:H7 Adherence Characteristics and Invasion Potential Reveals a New Attachment Pattern Shared by Intestinal E. coli. Infection and Immunity. 1995;63(5):2070–4.

18. Tzipori S, Gunzer F, Donnenberg MS, Montigny LD, Kaper JB, Donohue-Rolfe A. The Role of the eaeA Gene in Diarrhea and Neurological Complications in a Gnotobiotic Piglet Model of Enterohemorrhagic Escherichia coli Infection. Infection and Immunity. 1995;63:7.

19. Dean-Nystrom EA, Bosworth BT, Moon HW, O'Brien AD. Escherichia coli O157:H7 Requires Intimin for Enteropathogenicity in Calves. Infection and Immunity. 1998;66(0):4560–3.

20. Cornick N, Booher S, Moon H. Intimin Facilitates Colonization by Escherichia coli O157:H7 in Adult Ruminants. Infection and Immunity. 2002;70(5):2704–7.

21. Barnett Foster D. Modulation of the enterohemorrhagic *E. coli* virulence program through the human gastrointestinal tract. Virulence. 2013;4(4):315–23.

22. Hughes DT, Clarke MB, Yamamoto K, Rasko DA, Sperandio V. The QseC Adrenergic Signaling Cascade in Enterohemorrhagic E. coli (EHEC). PLoS Pathogens. 2009;5(8):e1000553.

23. Sperandio V, Torres AG, Jarvis B, Nataro JP, Kaper JB. Bacteria-host communication: The language of hormones. Proceedings of the National Academy of Sciences. 2003;100(15):8951–6.

24. Hirakawa H, Kodama T, Takumi-Kobayashi A, Honda T, Yamaguchi A. Secreted indole serves as a signal for expression of type III secretion system translocators in enterohaemorrhagic Escherichia coli O157:H7. Microbiology. 2009;155(2):541–50.

25. Kendall M. Interkingdom Chemical Signaling in Enterohemorrhagic Escherichia coli O157:H7. In: Lyte M, editor. Microbial Endocrinology: Interkingdom Signaling in Infectious Disease and Health. 2nd edition. Springer International Publishing. 2016. p.p. 201–213.

26. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature Methods. 2009;6(10):767–72.

27. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. Cell Host & Microbe. 2009;6(3):279–89.

28. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. Proceedings of the National Academy of Sciences. 2009;106(38):16422–7.

29. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Research. 2009;19(12):2308–16.

30. Gao B, Vorwerk H, Huber C, Lara-Tejero M, Mohr J, Goodman AL, et al. Metabolic and fitness determinants for in vitro growth and intestinal colonization of the bacterial pathogen Campylobacter jejuni. PLOS Biology. 2017;15(5):e2001390.

31. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, et al. Genome-wide identification of bacterial plant colonization genes. PLOS Biology. 2017;15(9):e2002860.

32. McCarthy AJ, Stabler RA, Taylor PW. Genome-Wide Identification by Transposon Insertion Sequencing of *Escherichia coli* K1 Genes Essential for *In Vitro* Growth, Gastrointestinal Colonizing Capacity, and Survival in Serum. Journal of Bacteriology. 2018;200(7).

33. de Moraes MH, Desai P, Porwollik S, Canals R, Perez DR, Chu W, et al. Salmonella Persistence in Tomatoes Requires a Distinct Set of Metabolic Functions Identified by Transposon Insertion Sequencing. Applied and Environmental Microbiology 2017;83(5).

34. Armbruster CE, Forsyth-DeOrnellas V, Johnson AO, Smith SN, Zhao L, Wu W, et al. Genome-wide transposon mutagenesis of Proteus mirabilis: Essential genes,

fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements. PLOS Pathogens. 2017;13(6):e1006434.

35. Capel E, Barnier J-P, Zomer AL, Bole-Feysot C, Nussbaumer T, Jamet A, et al. Peripheral blood vessels are a niche for blood-borne meningococci. Virulence. 2017;8(8):1808–19.

36. Zhu L, Charbonneau ARL, Waller AS, Olsen RJ, Beres SB, Musser JM. Novel Genes Required for the Fitness of Streptococcus pyogenes in Human Saliva. 2017;2(6):13.

37. George AS, Cox CE, Desai P, Porwolik S, Chu W, de Moraes MH, et al. Interactions of *Salmonella enterica* Serovar Typhimurium and *Pectobacterium carotovorum* within a Tomato Soft Rot. Applied and Environmental Microbiology. 2017;84(5).

38. White KM, Matthews MK, Hughes R, Sommer AJ, Griffitts JS, Newell PD, et al. A Metagenome-Wide Association Study and Arrayed Mutant Library Confirm *Acetobacter* Lipopolysaccharide Genes Are Necessary for Association with *Drosophila melanogaster*. G3: Genes, Genomes, Genetics. 2018;g3.300530.2017.

39. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature. 2018;557(7706):503–9.

40. Pritchard JR, Chao MC, Abel S, Davis BM, Baranowski C, Zhang YJ, et al. ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. PLoS Genetics. 2014;10(11):e1004782.

41. Chao MC, Abel S, Davis BM, Waldor MK. The design and analysis of transposon insertion sequencing experiments. Nature Reviews Microbiology. 2016;14(2):119–28.

42. Kimura S, Hubbard TP, Davis BM, Waldor MK. The Nucleoid Binding Protein H-NS Biases Genome-Wide Transposon Insertion Landscapes. mBio. 2016;7(4).

43. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. mBio. 2017;8(1).

44. Ritchie JM, Waldor MK. The Locus of Enterocyte Effacement-Encoded Effector Proteins All Promote Enterohemorrhagic Escherichia coli Pathogenicity in Infant Rabbits. Infection and Immunity. 2005;73(3):1466–74.

45. Hubbard TP, Chao MC, Abel S, Blondel CJ, Abel zur Wiesch P, Zhou X, et al. Genetic analysis of *Vibrio parahaemolyticus* intestinal colonization. Proceedings of the National Academy of Sciences. 2016;113(22):6283–8.

46. Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, et al. Comprehensive Assignment of Roles for Salmonella Typhimurium Genes in Intestinal Colonization of Food-Producing Animals. PLoS Genetics. 2013;9(4):e1003456.

47. Munoz-Lopez M, Garcia-Perez J. DNA Transposons: Nature and Applications in Genomics. Current Genomics. 2010;11(2):115–28.

48. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research. 2016;44(D1):D457–62.

49. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research. 2000;28(1):33–6.

50. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research. 2015;43(D1):D261–9.

51. Shields RC, Zeng L, Culp DJ, Burne RA. Genomewide Identification of Essential Genes and Fitness Determinants of *Streptococcus mutans* UA159. mSphere. 2018;3(1):e00031-18.

52. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Research. 2014;42(D1):D574–80.

53. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology. 2006;2:2006.0008.

54. Kato J, Hashimoto M. Construction of consecutive deletions of the Escherichia coli chromosome. Molecular Systems Biology. 2007;3:132.

55. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, et al. Update on the Keio collection of Escherichia coli single-gene deletion mutants. Molecular Systems Biology. 2009;5:335.

56. Sperandeo P, Pozzi C, Dehò G, Polissi A. Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the Escherichia coli yrbG–yhbG locus. Research in Microbiology. 2006;157(6):547–58.

57. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, et al. The Essential Genome of Escherichia coli K-12. 2018;9(1):18.

1130 58. Abel S, Waldor MK. Infant Rabbit Model for Diarrheal Diseases: Infant Rabbit
1131 Model for Diarrheal Diseases. Current Protocols in Microbiology. 2015;6A.6.1-
1132 6A.6.15.

1133 59. Ritchie JM, Rui H, Bronson RT, Waldor MK. Back to the Future: Studying Cholera
1134 Pathogenesis Using Infant Rabbits. mBio. 2010;1(1).

1135 60. Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. Analysis of Bottlenecks in
1136 Experimental Models of Infection. PLOS Pathogens. 2015;11(6):e1004823.

1137 61. Fu Y, Waldor MK, Mekalanos JJ. Tn-Seq Analysis of Vibrio cholerae Intestinal
1138 Colonization Reveals a Role for T6SS-Mediated Antibacterial Activity in the Host.
1139 Cell Host & Microbe. 2013;14(6):652–63.

1140 62. Abu Kwaik Y, Bumann D. Host Delivery of Favorite Meals for Intracellular
1141 Pathogens. PLOS Pathogens. 2015;11(6):e1004866.

1142 63. Meylan S, Porter CBM, Yang JH, Belenky P, Gutierrez A, Lobritz MA, et al.
1143 Carbon Sources Tune Antibiotic Susceptibility in Pseudomonas aeruginosa via
1144 Tricarboxylic Acid Cycle Control. Cell Chemical Biology. 2017;24(2):195–206.

1145 64. Murima P, McKinney JD, Pethe K. Targeting Bacterial Central Metabolism for
1146 Drug Development. Chemistry & Biology. 2014;21(11):1423–32.

1147 65. Ren W, Rajendran R, Zhao Y, Tan B, Wu G, Bazer FW, et al. Amino Acids As
1148 Mediators of Metabolic Cross Talk between Host and Pathogen. Frontiers in
1149 Immunology. 2018;9:319.

1150 66. Hubbard TP, D'Gama JD, Billings G, Davis BM, Waldor MK. Unsupervised
1151 Learning Approach for Comparing Multiple Transposon Insertion Sequencing
1152 Studies. 2019;4(1):13.

1153 67. Franzin FM, Sircili MP. Locus of Enterocyte Effacement: A Pathogenicity Island
1154 Involved in the Virulence of Enteropathogenic and Enterohemorragic Escherichia
1155 coli Subjected to a Complex Network of Gene Regulation. BioMed Research
1156 International. 2015;2015:1–10.

1157 68. Knutton S. A novel EspA-associated surface organelle of enteropathogenic
1158 Escherichia coli involved in protein translocation into epithelial cells. The EMBO
1159 Journal. 1998;17(8):2166–76.

1160 69. Kresse AU, Schulze K, Deibel C, Ebel F, Rohde M, Chakraborty T. Pas, a Novel
1161 Protein Required for Protein Secretion and Attaching and Effacing Activities of
1162 Enterohemorrhagic Escherichia coli. Journal of Bacteriology. 1998;180:10.

1163 70. Kresse AU, Rohde M. The EspD Protein of Enterohemorrhagic Escherichia coli Is
1164 Required for the Formation of Bacterial Surface Appendages and Is Incorporated

in the Cytoplasmic Membranes of Target Cells. Infection and Immunity. 1999;67:9.

71. Tatsuno I, Kimura H, Okutani A, Kanamaru K, Abe H, Nagai S, et al. Isolation and Characterization of Mini-Tn5Km2 Insertion Mutants of Enterohemorrhagic Escherichia coli O157:H7 Deficient in Adherence to Caco-2 Cells. Infection and Immunity. 2000;68(10):5943–52.

72. Sal-Man N, Setiaputra D, Scholz R, Deng W, Yu ACY, Strynadka NCJ, et al. EscE and EscG Are Cochaperones for the Type III Needle Protein EscF of Enteropathogenic Escherichia coli. Journal of Bacteriology. 2013;195(11):2481–9.

73. Gruenheid S, Sekirov I, Thomas NA, Deng W, O'Donnell P, Goode D, et al. Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic Escherichia coli O157:H7: Identification and characterization of NleA. Molecular Microbiology. 2004;51(5):1233–49.

74. Yen H, Sugimoto N, Tobe T. Enteropathogenic Escherichia coli Uses NleA to Inhibit NLRP3 Inflammasome Activation. PLOS Pathogens. 2015;11(9):e1005121.

75. Arbeloa A, Bulgin RR, MacKenzie G, Shaw RK, Pallen MJ, Crepin VF, et al. Subversion of actin dynamics by EspM effectors of attaching and effacing bacterial pathogens. Cellular Microbiology. 2008;10(7):1429–41.

76. Simovitch M, Sason H, Cohen S, Zahavi EE, Melamed-Book N, Weiss A, et al. EspM inhibits pedestal formation by enterohaemorrhagic *Escherichia coli* and enteropathogenic *E. coli* and disrupts the architecture of a polarized epithelial monolayer. Cellular Microbiology. 2010;12(4):489–505.

77. Munera D, Crepin VF, Marches O, Frankel G. N-Terminal Type III Secretion Signal of Enteropathogenic Escherichia coli Translocator Proteins. Journal of Bacteriology. 2010;192(13):3534–9.

78. Iyoda S, Watanabe H. ClpXP Protease Controls Expression of the Type III Protein Secretion System through Regulation of RpoS and GrlR Levels in Enterohemorrhagic Escherichia coli. Journal of Bacteriology. 2005;187(12):4086–94.

79. Tomoyasu T, Takaya A, Handa Y, Karata K, Yamamoto T. ClpXP controls the expression of LEE genes in enterohaemorrhagic Escherichia coli. FEMS Microbiology Letters. 2005;253(1):59–66.

80. MacRitchie DM, Ward JD, Nevesinjac AZ, Raivio TL. Activation of the Cpx Envelope Stress Response Down-Regulates Expression of Several Locus of Enterocyte Effacement-Encoded Genes in Enteropathogenic Escherichia coli. Infection and Immunity. 2008;76(4):1465–75.

1202 81. MacRitchie DM, Acosta N, Raivio TL. DegP Is Involved in Cpx-Mediated
1203 Posttranscriptional Regulation of the Type III Secretion Apparatus in
1204 Enteropathogenic Escherichia coli. Infection and Immunity. 2012;80(5):1766–72.

1205 82. Hara H, Yamamoto Y, Higashitani A, Suzuki H, Nishimura Y. Cloning, mapping,
1206 and characterization of the Escherichia coli prc gene, which is involved in C-
1207 terminal processing of penicillin-binding protein 3. Journal of Bacteriology.
1208 1991;173(15):4799–813.

1209 83. Kerr CH, Culham DE, Marom D, Wood JM. Salinity-Dependent Impacts of ProQ,
1210 Prc, and Spr Deficiencies on Escherichia coli Cell Structure. Journal of
1211 Bacteriology. 2014;196(6):1286–96.

1212 84. Pos KM. Drug transport mechanism of the AcrB efflux pump. Biochimica et
1213 Biophysica Acta (BBA) - Proteins and Proteomics. 2009;1794(5):782–93.

1214 85. Begley M, Gahan CGM, Hill C. The interaction between bacteria and bile. FEMS
1215 Microbiology Reviews. 2005;29(4):625–51.

1216 86. Christman MF, Storz G, Ames BN. OxyR, a positive regulator of hydrogen
1217 peroxide-inducible genes in Escherichia coli and Salmonella typhimurium, is
1218 homologous to a family of bacterial regulatory proteins. Proceedings of the
1219 National Academy of Sciences. 1989;86(10):3484–8.

1220 87. Daugherty A, Suvarnapunya AE, Runyen-Janecky L. The role of OxyR and
1221 SoxRS in oxidative stress survival in Shigella flexneri. Microbiological Research.
1222 2012;167(4):238–45.

1223 88. De Paepe M, Gaboriau-Routhiau V, Rainteau D, Rakotobe S, Taddei F, Cerf-
1224 Bensussan N. Trade-Off between Bile Resistance and Nutritional Competence
1225 Drives Escherichia coli Diversification in the Mouse Gut. PLoS Genetics.
1226 2011;7(6):e1002107.

1227 89. Berks BC. The Twin-Arginine Protein Translocation Pathway. Annual Review of
1228 Biochemistry. 2015;84(1):843–64.

1229 90. Green ER, Mecsas J. Bacterial Secretion Systems: An Overview. Microbiology
1230 Spectrum. 2016;4(1).

1231 91. Mickael CS, Lam P-KS, Berberov EM, Allan B, Potter AA, Koster W. Salmonella
1232 enterica Serovar Enteritidis tatB and tatC Mutants Are Impaired in Caco-2 Cell
1233 Invasion In Vitro and Show Reduced Systemic Spread in Chickens. Infection and
1234 Immunity. 2010;78(8):3493–505.

1235 92. Craig M, Sadik AY, Golubeva YA, Tidhar A, Slauch JM. Twin-arginine
1236 translocation system (tat) mutants of Salmonella are attenuated due to envelope
1237 defects, not respiratory defects: Role of Tat in Salmonella virulence. Molecular
1238 Microbiology. 2013;89(5):887–902.

93. Fujimoto M, Goto R, Hirota R, Ito M, Haneda T, Okada N, et al. Tat-exported peptidoglycan amidase-dependent cell division contributes to Salmonella Typhimurium fitness in the inflamed gut. PLOS Pathogens. 2018;14(10):e1007391.

94. Avican U, Doruk T, Östberg Y, Fahlgren A, Forsberg Å. The Tat Substrate SufI Is Critical for the Ability of Yersinia pseudotuberculosis To Cause Systemic Infection. Infection and Immunity. 2017;85(4).

95. Rajashekara G, Drozd M, Gangaiah D, Jeon B, Liu Z, Zhang Q. Functional Characterization of the Twin-Arginine Translocation System in Campylobacter jejuni. Foodborne Pathogens and Disease. 2009;6(8):935–45.

96. Lavander M, Ericsson SK, Broms JE, Forsberg A. The Twin Arginine Translocation System Is Essential for Virulence of Yersinia pseudotuberculosis. Infection and Immunity. 2006;74(3):1768–76.

97. Zhang L, Zhu Z, Jing H, Zhang J, Xiong Y, Yan M, et al. Pleiotropic effects of the twin-arginine translocation system on biofilm formation, colonization, and virulence in Vibrio cholerae. BMC Microbiology. 2009;9(1):114.

98. Pradel N, Ye C, Livrelli V, Xu J, Joly B, Wu L-F. Contribution of the Twin Arginine Translocation System to the Virulence of Enterohemorrhagic Escherichia coli O157:H7. Infection and Immunity. 2003;71(9):4908–16.

99. Samaluru H, SaiSree L, Reddy M. Role of SufI (FtsP) in Cell Division of Escherichia coli: Evidence for Its Involvement in Stabilizing the Assembly of the Divisome. Journal of Bacteriology. 2007;189(22):8044–52.

100. Tarry M, Arends SJR, Roversi P, Piette E, Sargent F, Berks BC, et al. The Escherichia coli Cell Division Protein and Model Tat Substrate SufI (FtsP) Localizes to the Septal Ring and Has a Multicopper Oxidase-Like Structure. Journal of Molecular Biology. 2009;386(2):504–19.

101. Stanley NR, Findlay K, Berks BC, Palmer T. Escherichia coli Strains Blocked in Tat-Dependent Protein Export Exhibit Pleiotropic Defects in the Cell Envelope. Journal of Bacteriology. 2001;183(1):139–44.

102. Bernhardt TG, De Boer PAJ. Screening for synthetic lethal mutants in Escherichia coli and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity: E. coli synthetic lethal screen. Molecular Microbiology. 2004;52(5):1255–69.

103. Fath MJ, Mahanty HK, Kolter R. Characterization of a purF operon mutation which affects colicin V production. Journal of Bacteriology. 1989;171(6):3158–61.

104. Hadjifrangiskou M, Gu AP, Pinkner JS, Kostakioti M, Zhang EW, Greene SE, et al. Transposon Mutagenesis Identifies Uropathogenic Escherichia coli Biofilm Factors. Journal of Bacteriology. 2012;194(22):6195–205.

105. Urdaneta V, Casadesús J. Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts. Frontiers in Medicine. 2017;4(163).

106. Shaffer CL, Zhang EW, Dudley AG, Dixon BREA, Guckes KR, Breland EJ, et al. Purine Biosynthesis Metabolically Constrains Intracellular Survival of Uropathogenic Escherichia coli. Infection and Immunity. 2017;85(1).

107. Kwon YM, Ricke SC, Mandal RK. Transposon sequencing: methods and expanding applications. Applied Microbiology and Biotechnology. 2016;100(1):31–43.

108. Barquist L, Boinett CJ, Cain AK. Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biology. 2013;10(7):1161–9.

109. Donnenberg MS, Kaper JB. Construction of an eae Deletion Mutant of Enteropathogenic Escherichia coli by Using a Positive-Selection Suicide Vector. Infection and Immunity. 1991;59:8.

110. Poteete AR, Rosadini C, St. Pierre C. Gentamicin and other cassettes for chromosomal gene replacement in *Escherichia coli*. BioTechniques. 2006;41(3):261–4.

111. Ferrieres L, Hemery G, Nham T, Guerout A-M, Mazel D, Beloin C, et al. Silent Mischief: Bacteriophage Mu Insertions Contaminate Products of Escherichia coli Random Mutagenesis Performed Using Suicidal Transposon Delivery Plasmids Mobilized by Broad-Host-Range RP4 Conjugative Machinery. Journal of Bacteriology. 2010;192(24):6418–27.

112. Cameron DE, Urbach JM, Mekalanos JJ. A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. Proceedings of the National Academy of Sciences. 2008;105(25):8736–41.

113. Chiang SL, Rubin EJ. Construction of a mariner -based transposon for epitope-tagging and genomic targeting. Gene. 2002;296(1–2):179–85.

114. Hubbard TP, Billings G, Dörr T, Sit B, Warr AR, Kuehl CJ, et al. A live vaccine rapidly protects against cholera in an infant rabbit model. Science Translational Medicine. 2018;10(445):eaap8423.

115. Abel S, Abel zur Wiesch P, Chang H-H, Davis BM, Lipsitch M, Waldor MK. Sequence tag–based analysis of microbial population dynamics. Nature Methods. 2015;12(3):223–6.

116. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software. 2010;36(3).

117. Collington GK, Booth IW, Knutton S. Rapid modulation of electrolyte transport in Caco-2 cell monolayers by enteropathogenic Escherichia coli (EPEC) infection. Gut. 1998;42(2):200–7.

118. Stincone A, Daudi N, Rahman AS, Antczak P, Henderson I, Cole J, et al. A systems biology approach sheds new light on Escherichia coli acid resistance. Nucleic Acids Research. 2011;39(17):7512–28.

119. Dörr T, Möll A, Chao MC, Cava F, Lam H, Davis BM, et al. Differential Requirement for PBP1a and PBP1b in In Vivo and In Vitro Fitness of Vibrio cholerae. Infection and Immunity. 2014;82(5):2115–24.

120. Cremers CM, Knoefler D, Vitvitsky V, Banerjee R, Jakob U. Bile salts act as effective protein-unfolding agents and instigators of disulfide stress in vivo. Proceedings of the National Academy of Sciences. 2014;111(16):E1610–9.

121. Käll L, Krogh A, Sonnhammer EL. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biology. 2004;338(5):1027–36.

122. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. Journal of Molecular Biology. 2018;430(15):2237–43.

123. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 2015;10(6):845–58.

124. Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics. 2014;30(6):884–6.

125. Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD. Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics. 2010;26(22):2811–7.

Figure Legends

Figure 1: Analysis of essential genes in EHEC EDL933 and comparison to K-12 MG1655.

A) Distribution of percentage TA site disruption for all genes in EHEC EDL933. Genes are classified by EL-ARTIST as either underrepresented (red), regional (purple), or neutral (blue).

B) Transposon-insertion profiles of representative underrepresented, regional, and neutral genes.

C) GC content (%) of EDL933 genes with and without homologs in K-12 MG1655, classified by TIS classification (neutral or non-neutral). Neutral and non-neutral genes within each gene type (divergent or homolog) are compared using a Mann-Whitney U test with a Bonferroni correction. Ratios of neutral to non-neutral genes for each gene type are compared using a Fisher's Exact Test; (**) indicates a p-value of <0.01 and (***) indicate a p-value of <0.001.

D) Distribution of percentage TA site disruption for all genes in K-12 MG1655. Genes are classified by EL-ARTIST as underrepresented (red), regional (purple), or neutral (blue) using the same parameters as for the EDL933 library.

E) Genes classified as non-neutral in the ELD933 TIS library were compared to the MG1655 TIS library and categorized as either lacking a homolog (green), having the same classification in both libraries (red), or being non-neutral in EDL933 and neutral in MG1655 (blue).

F) KEGG pathway information for genes that are non-neutral in EDL933 and neutral in MG1655.

Figure 2: Identification of EHEC genes required for intestinal colonization.

A) Distribution of percentage TA site disruption for all genes in EDL933 in the library used to inoculate infant rabbits (top) and in a representative library recovered from a rabbit colon two days after infection (bottom).

B) Schematic of Con-ARTIST classification scheme. Con-ARTIST utilizes iterative resampling of the inoculum data set to generate 100 simulated control data sets and compares relative abundance of mutants in these simulated control data sets relative to the passaged library. Genes with sufficient data (≥ 5 TA sites disrupted) are then classified based on a dual standard of attenuation (mean $\log_2$ fold change ≤ -2) and consistency (Mann Whitney U p-value <0.05) as either queried (blue) or conditionally depleted (red).

C) Distribution of percentage TA site disruption for all genes in the inoculum library (top) and a representative library recovered from the rabbit colon (bottom) overlaid with the

classifications described in panel B. Genes with insufficient data are removed from the bottom panel.

D) Distribution of Con-ARTIST gene classifications (insufficient data (ID, black), queried (Q, blue), conditionally depleted (CD, red)) in each library recovered from seven infant rabbit colons two days post infection as compared to the inoculum.

E) Conditionally depleted genes (defined by Con-ARTIST consensus approach) by Clusters of Orthologous Groups (COG) classification. COG enrichment index (displayed as log2 enrichment) is calculated as the percentage of the CD genes assigned to a specific COG divided by the percentage of genes in that COG in the entire genome. A two-tailed Fisher's exact test with a Bonferroni correction was used to test the null hypothesis that enrichment is independent of TIS classification. (***), p-value <0.0001.

F) KEGG pathways of EHEC genes classified as conditionally depleted by Con-ARTIST consensus approach.

G) Distribution of PC1 scores across all EHEC genes. Red bins fall within the lowest 10% of PC1 scores.

Figure 3: Con-ARTIST and CompTIS-based classification of LEE genes and T3SS effectors.
A) Artemis plots of reads in the LEE pathogenicity island in the control-simulated inoculum library (top) and a representative library recovered from the rabbit colon (middle). The genes in the LEE are displayed at the bottom. The color of the gene corresponds to its classification. Maroon genes were categorized as conditionally depleted by Con-ARTIST consensus and fell in the bottom 10% of glPC1 scores by CompTIS; red genes had a glPC1 score in the bottom 10% of the distribution, but not classified as CD by Con-ARTIST. Gray genes did not meet the glPC1 cutoff and were not classified by Con-ARTIST.

B) Schematic showing classification of the LEE genes and non-LEE-encoded effectors. Color symbols as above. Orange indicates the gene was identified as CD by Con-ARTIST and had a glPC1 score above 10%. LEE-encoded effectors, non-LEE encoded effectors, chaperones/substrate selection proteins, and regulators are indicated in boxes.

Figure 4: Validation of colonization defects in selected mutants.
A) Competitive indices of indicated mutants vs wild type EHEC. Bar-coded mutants were co-inoculated with bar-coded wild-type EHEC into infant rabbits and recovered two days later from the colons of infected rabbits. Relative abundance of each mutant was determined by sequencing the barcodes. (**) p-values < 0.01 and (***) p-value <0.001. Δ*hupB,* which had a glPC1 score in the bottom 10%, but was not classified as CD by Con-ARTIST, is highlighted in blue.

Figure 5: Effector translocation and survival in response to various gastro-intestinal stressors by mutants defective in colonization.

A) Normalized effector translocation of mutants compared to WT. Δ*escN,* a mutant that abrogates T3SS activity, was used as a control. Mutants were tested for their ability to translocate EspF-TEM1 into HeLa cells, as measured by a shift in emission spectra from 520 to 450 nm. Fluorescence was normalized to WT levels. Geometric means and geometric standard deviations are plotted.

B) MIC for NaCl (osmotic stress) and crude bile for the indicated mutants. Bold text highlights values differing from the wild-type.

C) Normalized acid resistance. Mutants were tested for their ability to survive low acid shock. Survival is shown as a percentage of the acid resistance of the WT. Geometric mean and geometric standard deviation are plotted.

Figure 6: CvpA promotes EHEC resistance to deoxycholate.

A) Dilution series of WT, Δ*cvpA* mutant, and Δ*cvpA* mutant with arabinose-inducible *cvpA* complementation plasmid plated on LB, LB 1% deoxycholate (DOC), LB 1% cholate (CHO), or LB 1% DOC + 0.2% arabinose.

B) Optical density of WT and Δ*cvpA* grown in LB and two concentrations of DOC, added at the indicated arrow. The average of three readings is plotted with errors bars indicate standard deviation.

C) MIC of antimicrobial compounds for WT and Δ*cvpA* and Δ*acrAB* mutants. Units are mg/mL unless specified otherwise. Bolded values are those different than the wild-type.

D) Predicted CvpA topology diagram.

Supplementary Figure Legends

S1: Sequencing saturation of TIS libraries

Reads were randomly sampled from each library and the percentage of TA sites disrupted in each randomly selected pool were plotted for the EDL933 library (A), MG1655 library (B), the inoculum library used to infect infant rabbits (C), and the libraries recovered from 7 rabbit colons (D-J).

S2: Assessment of non-neutral EHEC EDL933 genes.

A) Non-neutral genes (defined by EL-ARTIST as either regional or underrepresented) by Clusters of Orthologous Groups (COG) classification. COG enrichment index (displayed as log2 enrichment) is calculated as defined in (51) as the percentage of the CD genes assigned to a specific COG divided by the percentage of genes in that COG in the entire genome. A two-tailed Fisher's exact test with a Bonferroni correction was

used to test the null hypothesis that enrichment is independent of TIS status. p-values considered to be significant if <0.002. Single asterisks (*) indicates p-value <0.002, double asterisks (**) indicate p-value <0.001, and triple asterisks (***) indicate p-value <0.0001.

B) GC content (%) of EDL933 genes classified as either neutral (blue) or non-neutral (regional + underrepresented; red) by TIS. Distributions are compared using a Mann-Whitney U non-parametric test; (***) p-value of <0.0001.

C) GC content (%) of EDL933 genes classified as either having homologs in MG1655 (homolog) or lacking homologs (divergent). Distributions are compared using a Mann-Whitney U test; (***)p-value of <0.0001.

D) TA insertions across *kdsC* in EDL933 (left) and MG1655 (right).

S3: Con-ARTIST and CompTIS classification of genes important for colonization

A-G) Distribution of percentage TA site disruption in libraries recovered from 7 rabbit colons. These distributions are overlaid with Con-ARTIST classification (queried, blue; CD (conditionally depleted), red) as described in Figure 2B.

H) CD genes were grouped by consensus across animals. Many genes are CD in only one rabbit; fewer are classified as CD across all 7 animals. A standard of consensus of 5 or more animals was chosen to determine the list of CD genes, indicated with an asterisk (*).

I) Variance explained by each gene-level (gl) principal component for glPCA performed across the 7 rabbit screens.

J) Gene-level principal component 1 (glPC1) coefficients for each rabbit dataset.

H) Heatmap of the log2 fold change for each gene with a glPC1 score that falls within the bottom 10% of the distribution. Each column represents genes from a separate rabbit replicate. Genes are ordered by PC1 score, lowest at the top of the heatmap and highest at the bottom.

S4: In vitro growth and morphology of mutants.

A) 17 mutant strains plus the wild-type were grown in LB and turbidity measured by optical density. The average of three readings with the standard deviation is plotted.

B) Cell-shape defects of Δ*sufI,* Δ*envC,* Δ*tatABC*, and Δ*prc* mutants in high osmolality media. Morphology in LB (top) or LB supplemented with 0.3M NaCl (bottom) is shown.

S5: Characterization of Δ*cvpA*.

A) Biofilm production in WT and Δ*cvpA* using crystal violet staining and absorption. Levels were normalized to a percent of the WT value; three samples were analyzed and the geometric means and geometric standard deviation are plotted. The differences between the two groups were not significant (n.s.) by Mann-Whitney U.

B) WT and Δ*cvpA* struck to single colonies on an agar plate made with YESCA media supplemented with Congo Red to detect curli fibers.

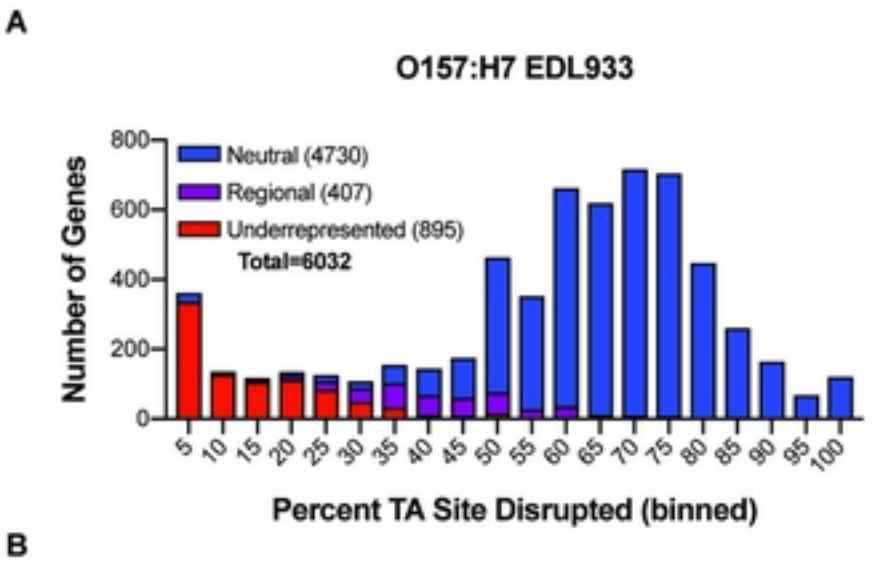
C) WT and Δ*cvpA* struck to single colonies on an agar plate containing minimal media with no exogenous purines.

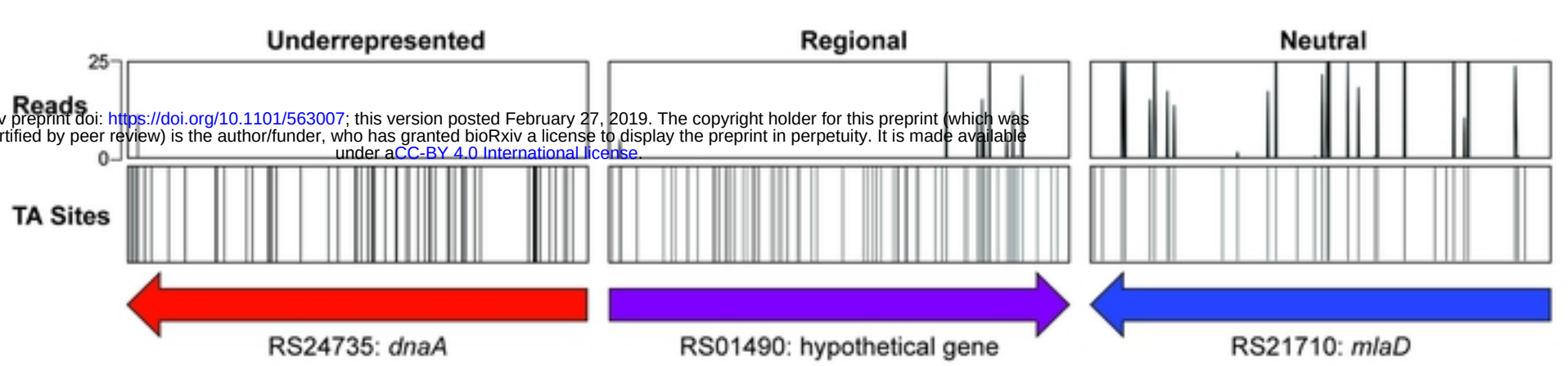
D) Dilution series of *Vibrio cholerae* C6706 WT and *cvpA*::tn plated on LB and LB 1% deoxycholate (DOC).

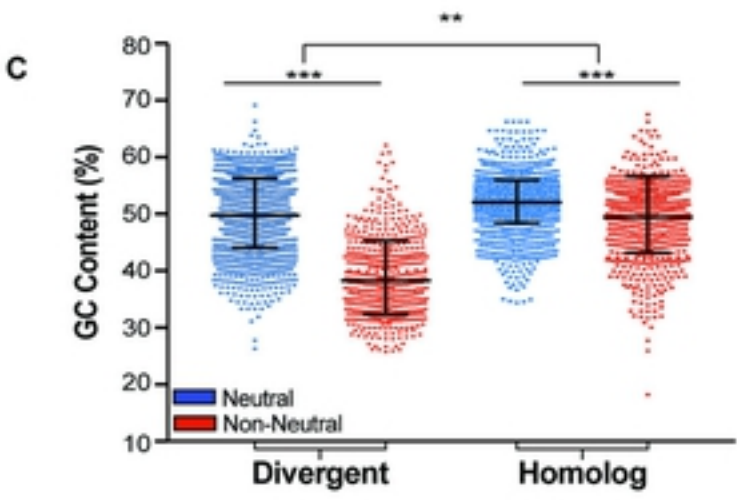
E) Dilution series of *Vibrio parahaemolyticus* WT and Δ*cvpA* plated on LB and LB 1% deoxycholate (DOC).

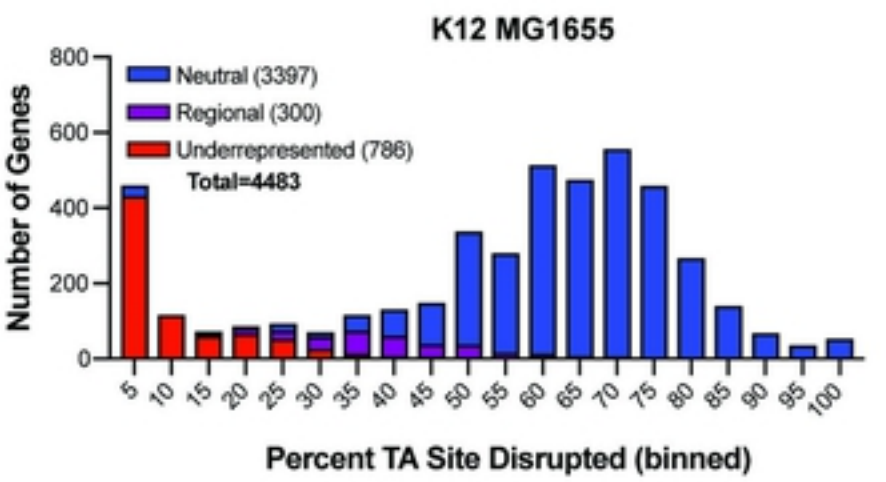
Supplementary Tables Captions:

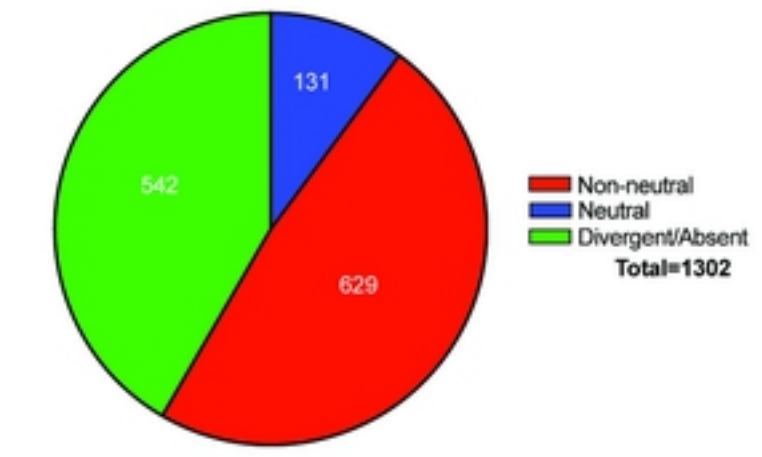
S1) RS to Z Annotation
S2) EHEC EL-ARTIST
S3) EHEC Non-Neutral KEGG
S4) K12 EL-ARTIST
S5) EHEC Unique Non-Neutral
S6) Con-ARTIST and CompTIS
S7) EHEC CD Genes KEGG
S8) Strains
S9) Oligos







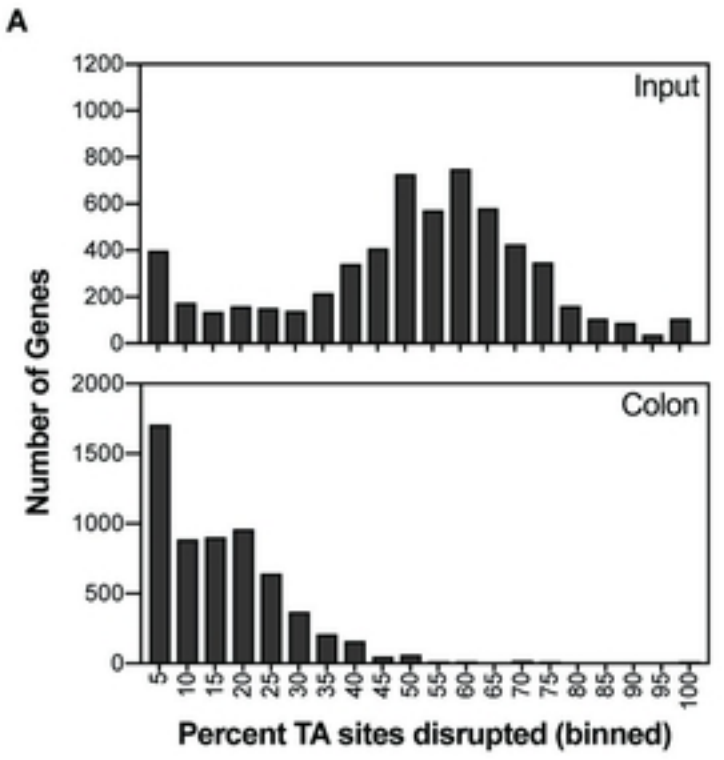


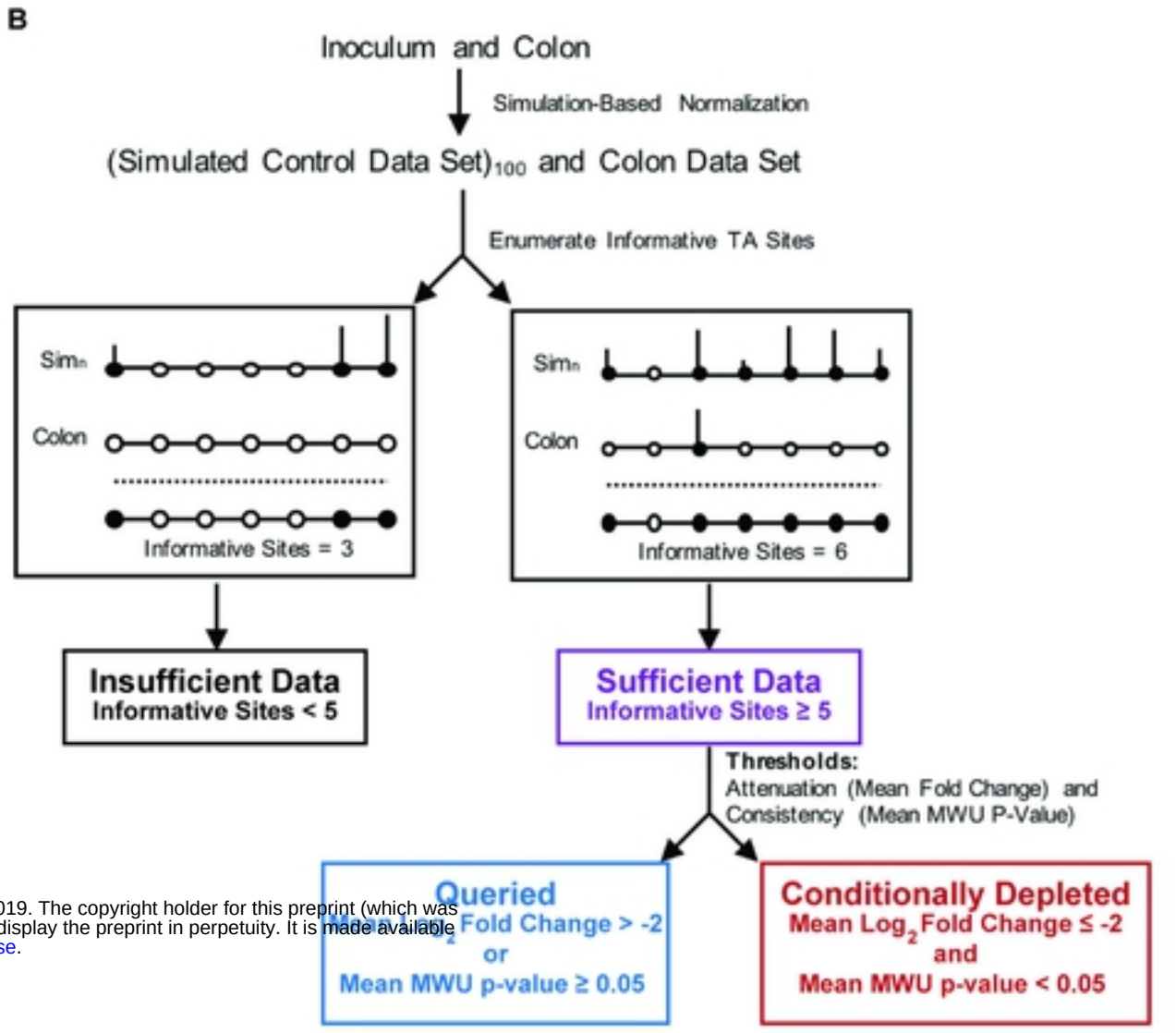


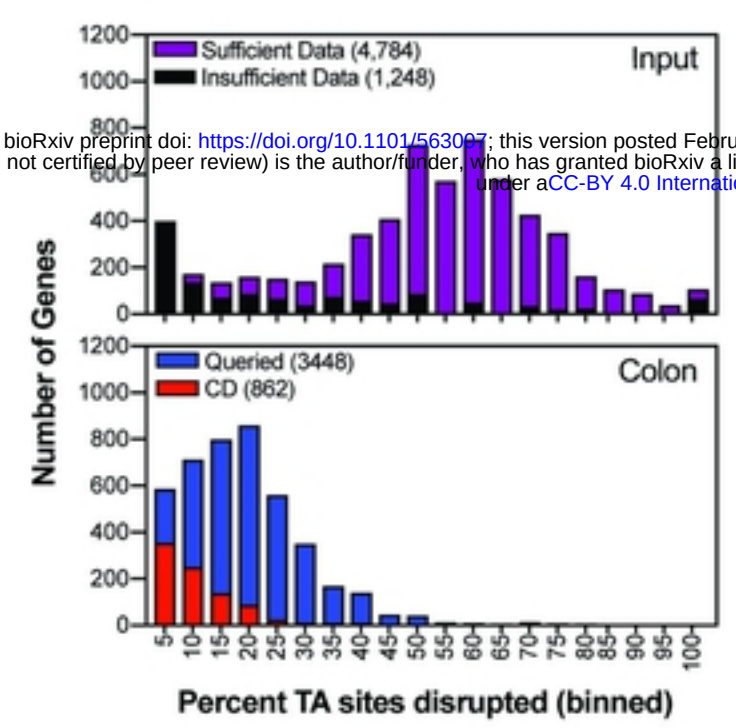
F

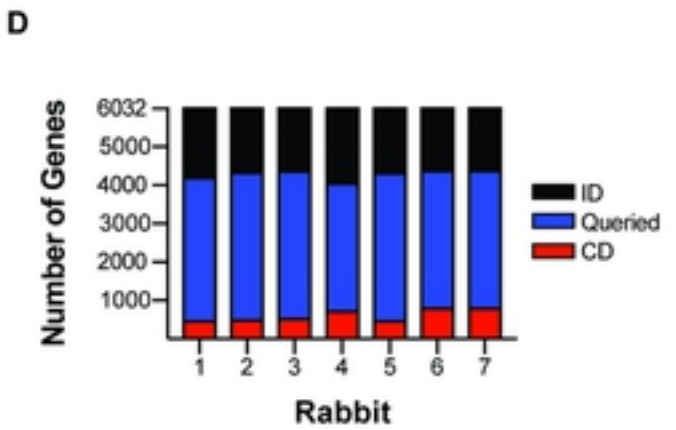
KEGG Pathway	Number of Genes
Metabolic pathways	16
Two-component system	5
Microbial metabolism in diverse environments	5
Galactose metabolism	4
Glycerophospholipid metabolism	3
Aminoacyl-tRNA biosynthesis	3
Quorum sensing	3
Pyrimidine metabolism	3
Phosphotransferase system (PTS)	3
ABC transporters	3

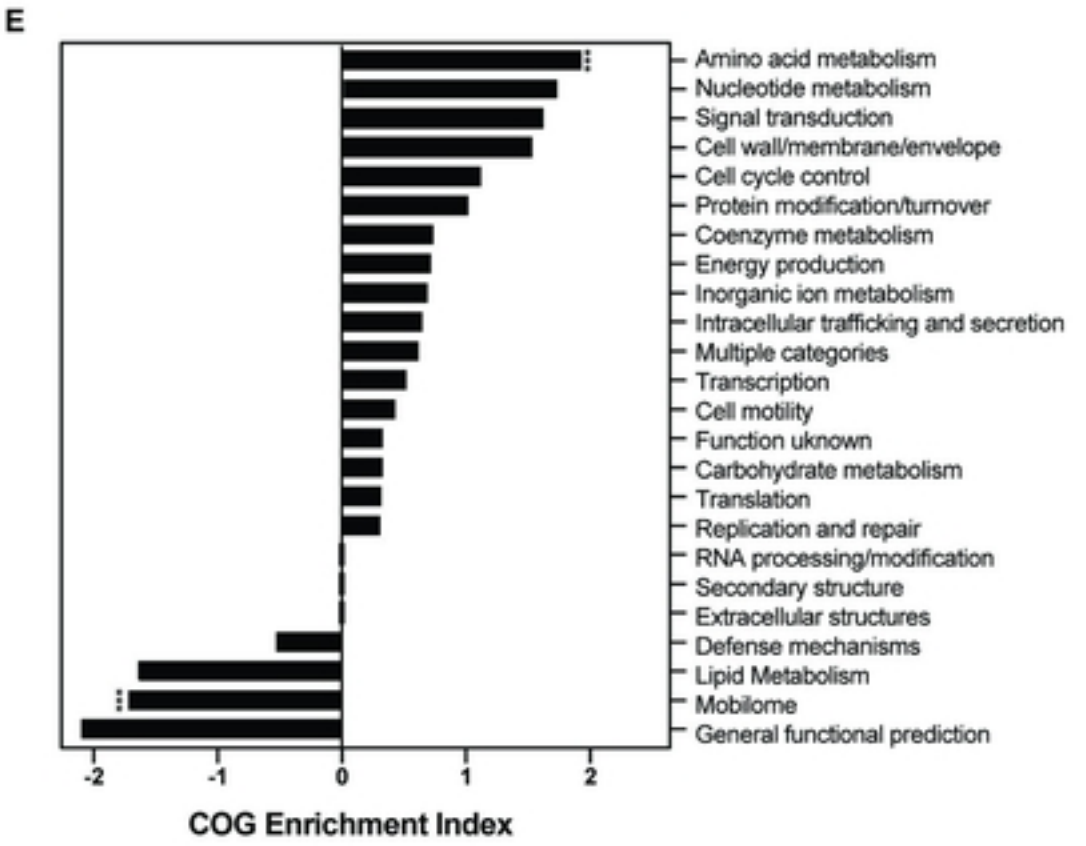
Figure 1











F

KEGG Pathway	Number of Genes
Metabolic pathways	102
Biosynthesis of secondary metabolites	61
Biosynthesis of antibiotics	52
Biosynthesis of amino acids	44
Microbial metabolism in diverse environments	30
Carbon metabolism	19
Two-component system	17
2-oxocarboxylic acid metabolism	12
Glycine, serine and threonine metabolism	10
Purine metabolism	10

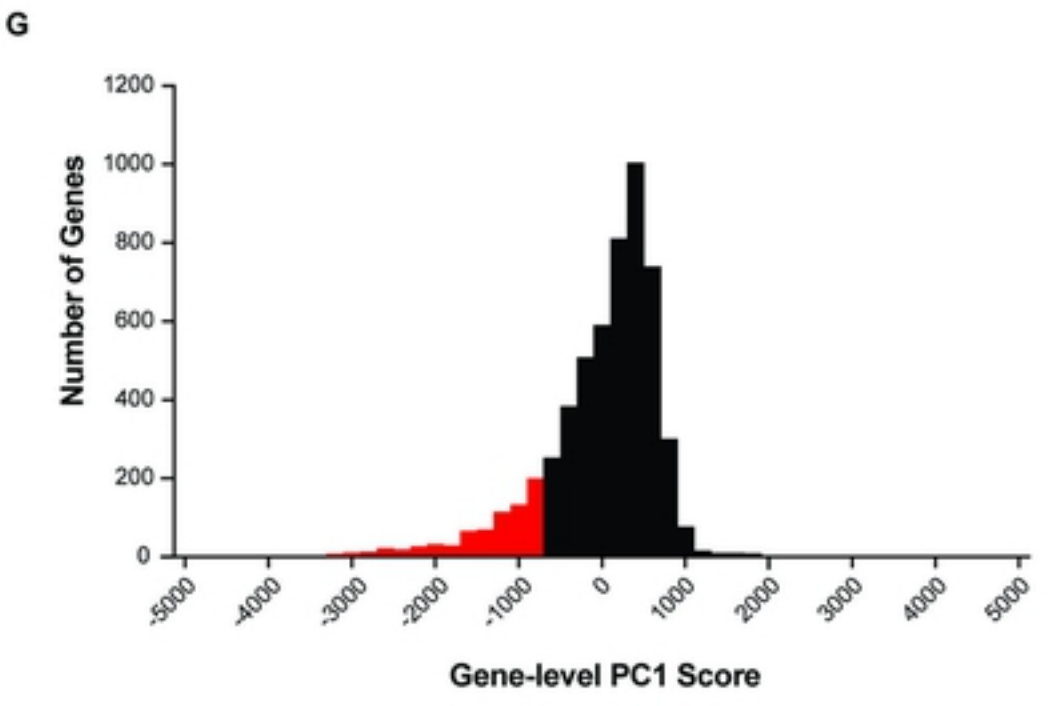


Figure 2

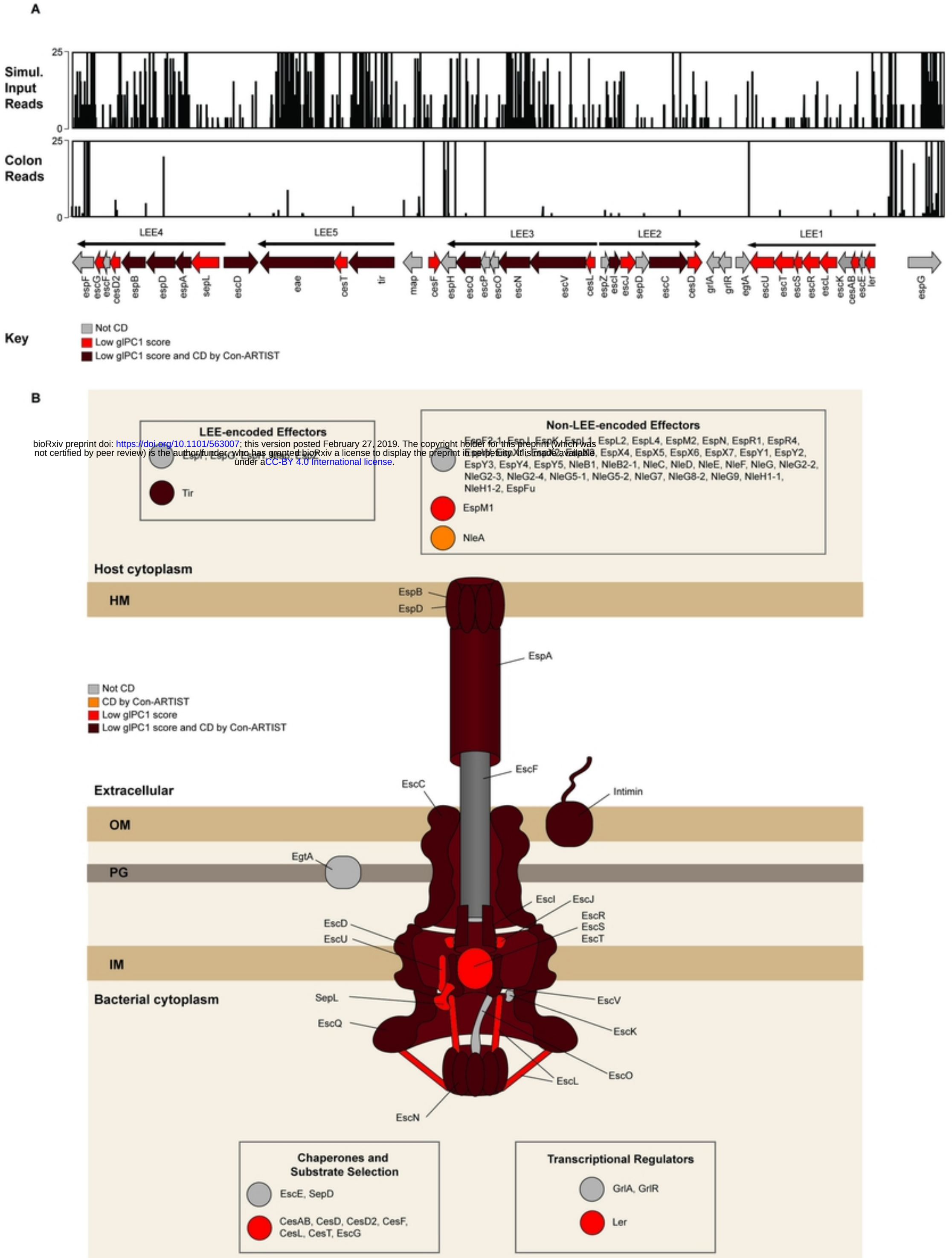


Figure 3

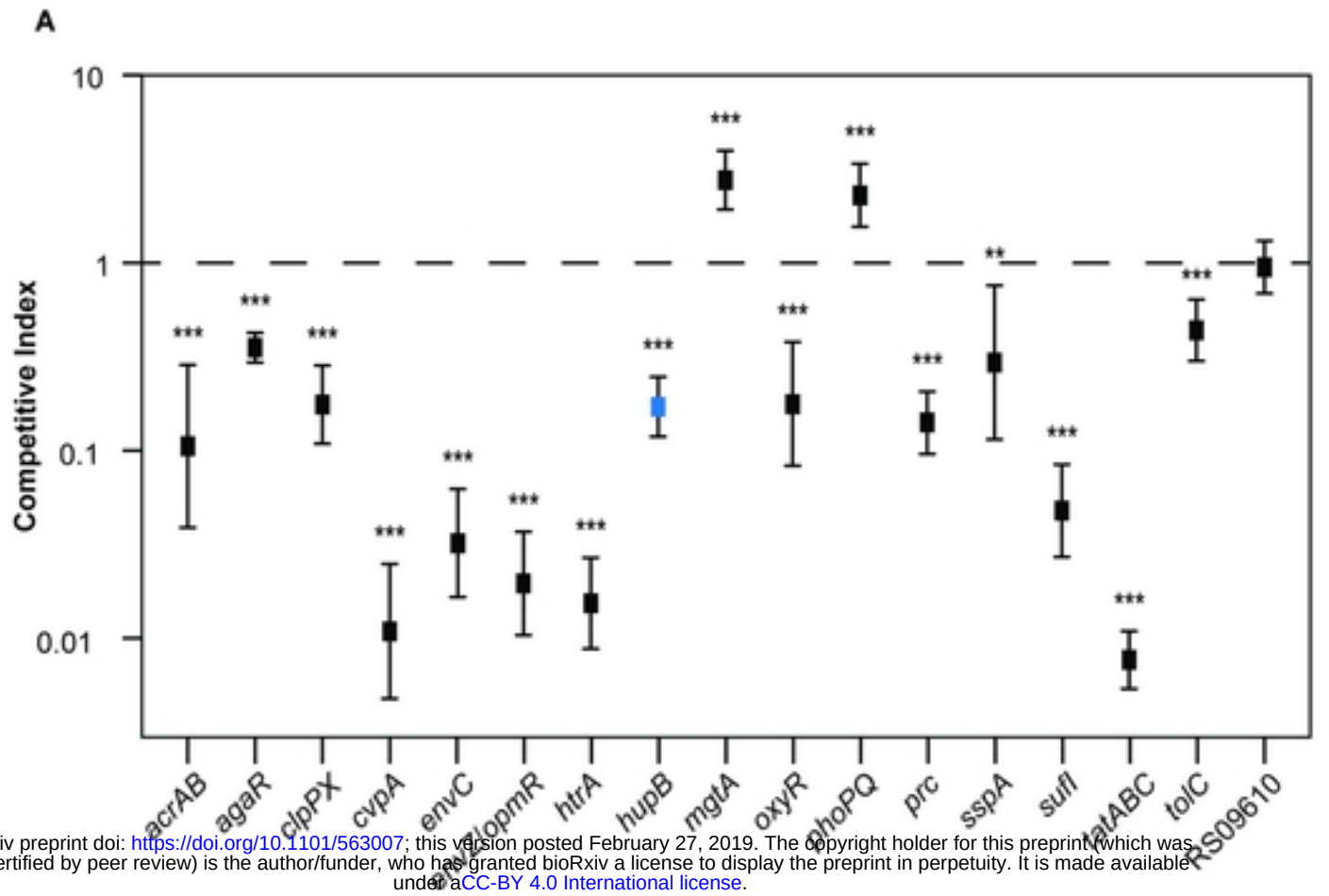
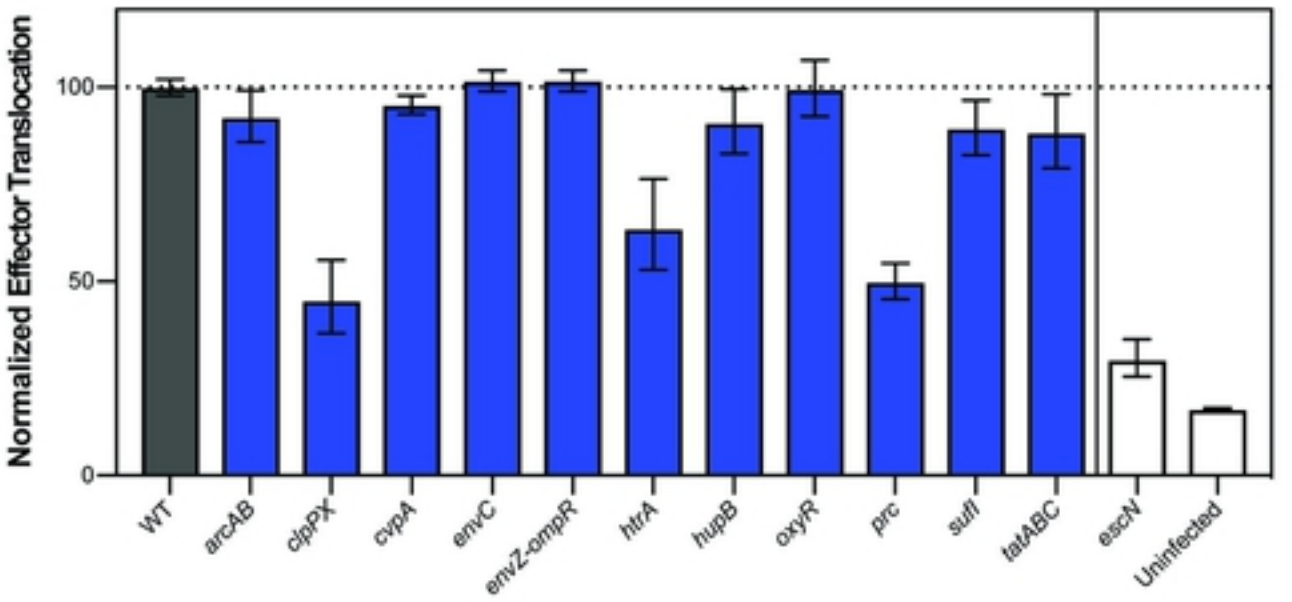


Figure 4

A



B

Strain	NaCl (mM)	Bile (%)
WT	1500	12
Δ*acrAB*	1500	**0.05**
Δ*clpPX*	1500	**6**
Δ*cvpA*	1500	**3**
Δ*envC*	**750**	**0.05**
Δ*envZ-ompR*	**750**	**1.5**
Δ*htrA*	1500	12
Δ*hupB*	1500	12
Δ*oxyR*	1500	**0.75**
Δ*prc*	**750**	**6**
Δ*sufI*	1500	12
Δ*tatABC*	**750**	**6**

C

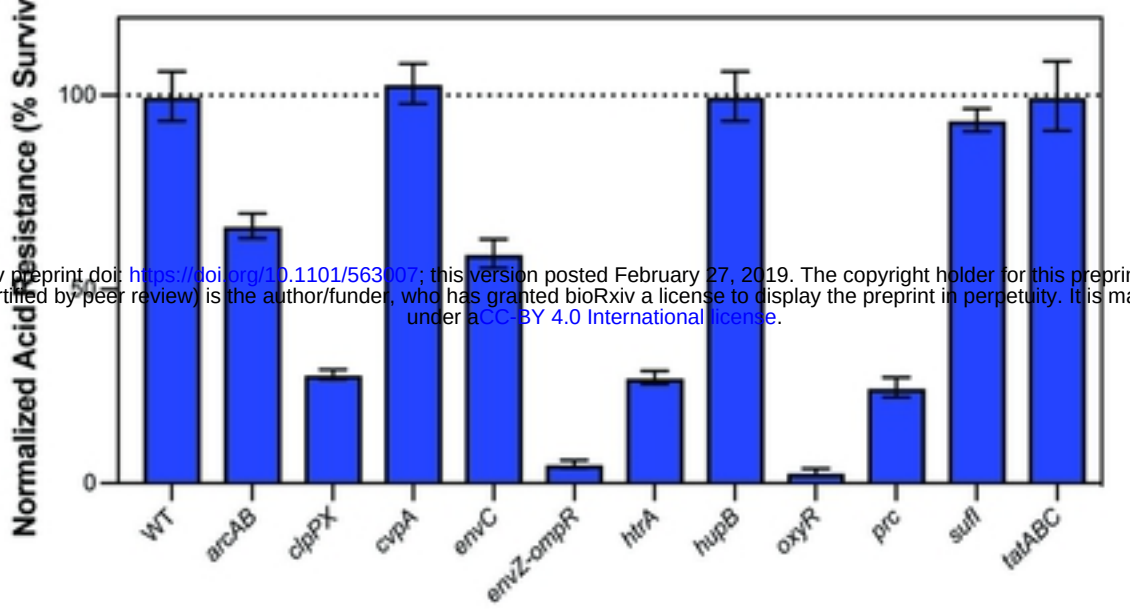
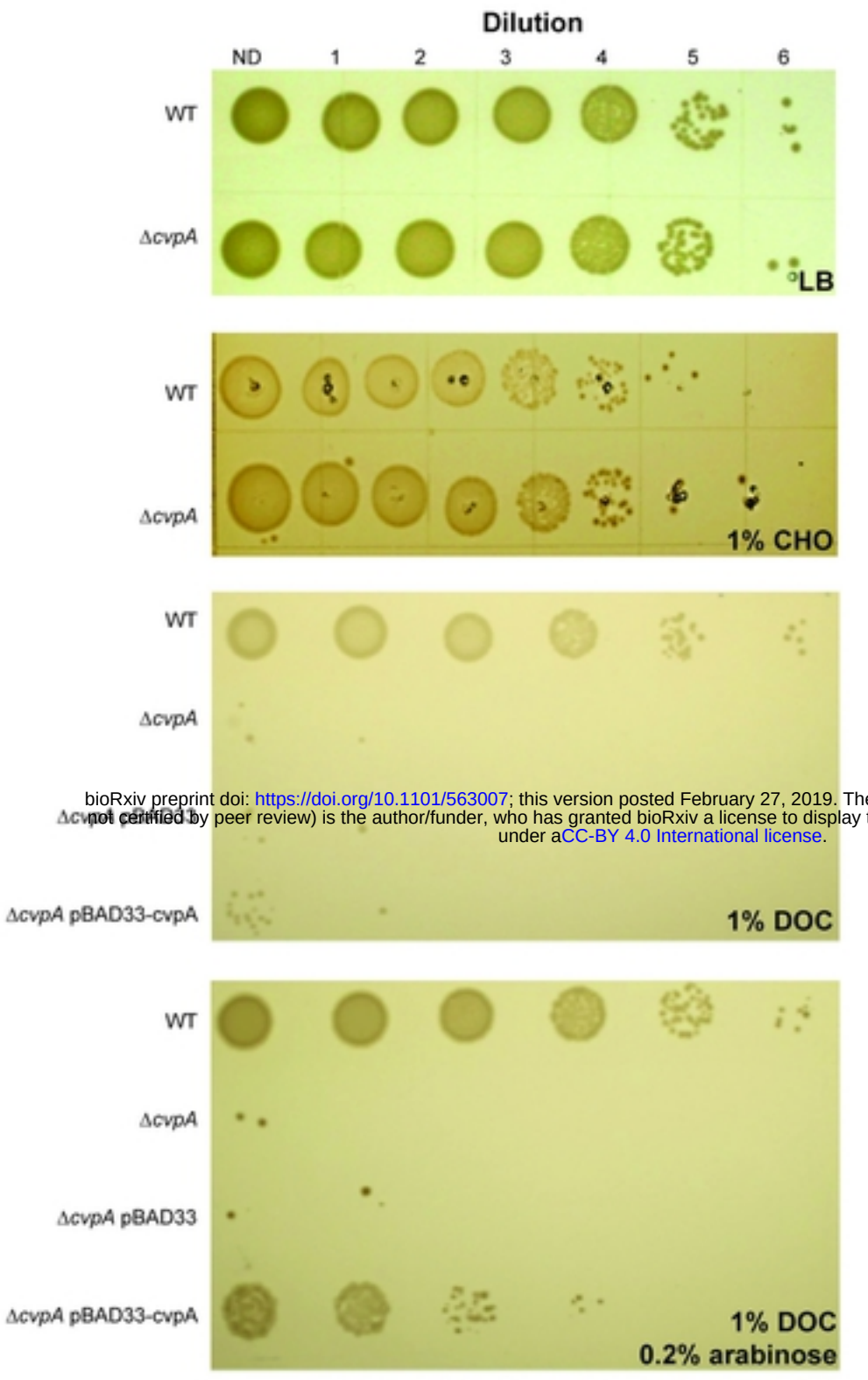
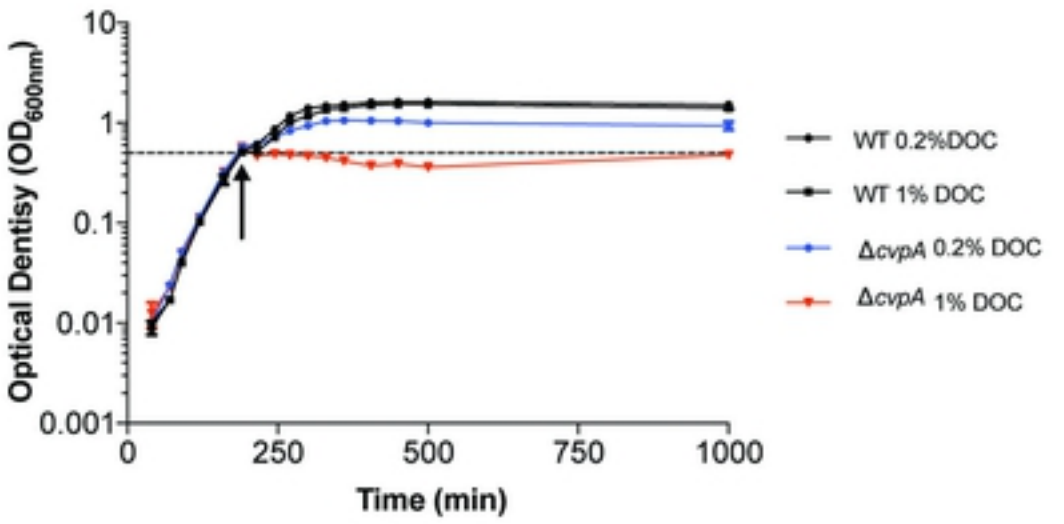


Figure 5

A



B



C

Antimicrobial Compound	WT	*ΔcvpA*	*ΔacrAB*
Deoxycholate (%)	2.5	**0.08**	**0.04**
Cholate (%)	2.5	**1.25**	**0.16**
Acridine Orange	>1000	>1000	**250**
Benzalkonium Chloride	6.25	6.25	**1.56**
Chloramphenicol	2.5	2.5	**0.25**
Ethidium Bromide	250	250	**7.81**
Penicillin G	6.25	6.25	**1.56**
Piperacillin	0.31	0.31	**0.02**
Polymyxin B	1.17	1.17	**0.59**
Nalidixic Acid	2.5	2.5	**0.31**
SDS	5	5	**0**

D

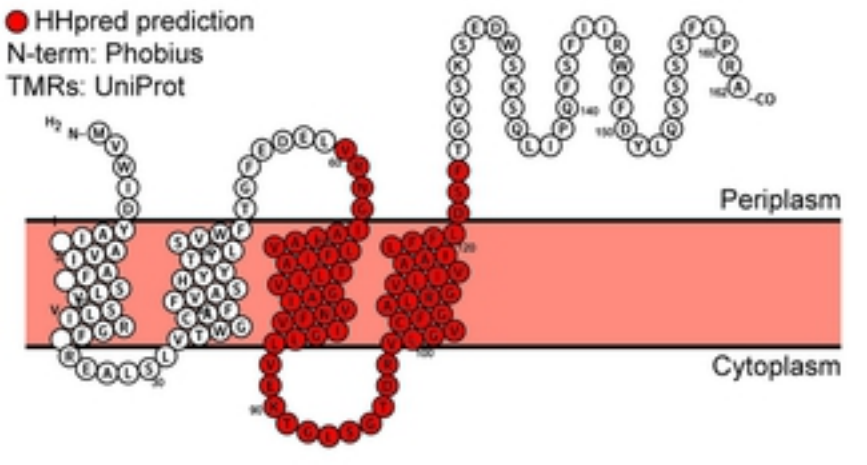


Figure 6