

1 **The highly evolvable nature of the antibiotic efflux protein TolC limits use of**  
2 **phages and bacterial toxins as antibacterial agents**

3 Yusuf Talha Tamer<sup>1</sup>, Ilona Gaszek<sup>1</sup>, Marinelle Rodrigues<sup>1</sup>, Fatma Sevde Coskun<sup>2</sup>, Michael Farid<sup>3</sup>,  
4 Andrew Y. Koh<sup>4,5</sup>, William Russ<sup>1,6</sup>, Erdal Toprak<sup>1,7,\*</sup>

5 <sup>1</sup> Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, Texas,  
6 USA

7 <sup>2</sup> Department of Immunology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

8 <sup>3</sup> Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States  
9 of America

10 <sup>4</sup> Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas, United  
11 States of America

12 <sup>5</sup> Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas, United  
13 States of America

14 <sup>6</sup> Center for Alzheimer's and Neurodegenerative Diseases, University of Texas Southwestern Medical  
15 Center, Dallas, Texas, USA

16 <sup>7</sup> Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

17 \* Corresponding Author

18 **Abstract**

19 **Bacteriophages and bacterial toxins are promising antibacterial agents to treat**  
20 **infections caused by multidrug resistant (MDR) bacteria. In fact, bacteriophages**  
21 **have recently been successfully used to treat life-threatening infections caused**  
22 **by MDR bacteria [1-3]. One potential problem with using these antibacterial**  
23 **agents is the evolution of resistance against them in the long term. Here, we**  
24 **studied the fitness landscape of the *Escherichia coli* TolC protein, an outer**  
25 **membrane protein that is exploited by a pore forming toxin called colicin E1 and**  
26 **by TLS-phage [4, 5]. By systematically assessing the distribution of fitness**  
27 **effects (DFEs) of ~9,000 single amino acid replacements in TolC using either**  
28 **positive (antibiotics and bile salts) or negative (colicin E1 and TLS-phage)**  
29 **selection pressures, we quantified evolvability of the TolC. We demonstrated that**  
30 **the TolC is highly optimized for the efflux of antibiotics and bile salts. In contrast,**

31 **under colicin E1 and TLS phage selection, TolC sequence is very sensitive to**  
32 **mutation. Our findings suggest that TolC is a highly evolvable target limiting the**  
33 **potential clinical use of bacteriophages and bacterial toxins.**

34

35 **Main**

36 TolC is an outer membrane protein conserved across Gram-negative bacteria and  
37 critical for the protection of bacterial cells against the toxicity of antimicrobial  
38 compounds such as antibiotics and bile salts (Figure 1A) [6-9]. TolC forms a  
39 homotrimeric channel, composed of a  $\beta$ -barrel domain that spans the outer membrane  
40 and a helical coiled-coil bundle that extends ~10 nm into the periplasmic space. TolC  
41 can partner with the AcrA-AcrB, MacA-MacB, and EmrA-EmrB protein pairs to form  
42 different efflux pump complexes in *E. coli* [10]. These efflux pumps render *E. coli*  
43 intrinsically resistant to several antibiotics such as  $\beta$ -lactams and macrolides [10, 11].  
44 TolC also effluxes bile salts which are abundant in the human gut (Figure 1A) [12, 13].  
45 Therefore, although *tolC* is typically not considered an essential gene for *E. coli*, the  
46 loss of *tolC* is costly in the presence of antibiotics or bile salts (Figure S1, S2).  
47 Interestingly, as TolC is exploited by both colicin E1 and the lytic TLS bacteriophage as  
48 a receptor, exposure to antibiotics (or bile salts) or colicin E1 (or TLS phage) creates  
49 opposing selective forces on maintaining the function of TolC, creating a convoluted  
50 fitness landscape where the evolutionary dynamics of TolC is highly unpredictable.

51 Mutational robustness is a common characteristic of evolved. For evolutionary  
52 success, a protein must tolerate spontaneous mutations for both survival and  
53 functional innovation. In the past decade, there have been many studies  
54 systematically assessing the fitness effects of mutations, particularly single amino acid  
55 replacements, on the function of proteins. Several of these studies have shown that  
56 protein function is robust to most single amino acid replacements [14-16]. In other  
57 cases, however, it was shown that many mutations can significantly deteriorate or  
58 even impair protein function [14-16]. As was originally proposed by Fisher, this  
59 problem gets even more complex because of the pleiotropic effects of mutations that  
60 can improve or worsen multiple traits simultaneously [17]. To date, most studies of the

61 evolutionary dynamics of proteins have focused on the selection pressure imposed by a  
62 single, specific growth condition, although the natural process often involves multiple,  
63 potentially opposing selection pressures [18-20]. This difference limits our  
64 understanding of how functional protein sequences have adapted to environmental  
65 fluctuations.

66 According to Fisher's fundamental theorem, the distribution of fitness effects (DFE) for  
67 mutations can be used as a metric for protein evolvability [17]. Using this theorem, we  
68 have previously shown that the width of the DFE is a good predictor for the rate of  
69 evolution of antibiotic resistance [21]. If a biological system is highly fit and robust to  
70 genetic perturbations, mutations are expected to have small fitness effects yielding a  
71 narrow DFE, centered around neutrality (Figure 1B, point C). However, if the biological  
72 system has low fitness and is sensitive to genetic perturbations, mutations are  
73 expected to have larger fitness effects and hence the DFE will be wider (Figure 1B,  
74 points A and B). Of note, in Figure 1B, we use normal distributions with varying widths  
75 to represent the DFEs, but realistically there is no way of predicting the shapes of  
76 these distributions and there may be outlier groups of beneficial or deleterious  
77 mutations.

78 In this study, we explore the evolvability of the efflux protein TolC using a saturation  
79 mutagenesis library which contains all possible amino acid replacements for each  
80 position of the TolC protein (barring the start codon). Utilizing both positive and  
81 negative selection, we performed a deep-sequencing based fitness assay to quantify  
82 the fitness landscape of TolC. We **systematically assessed the distribution of fitness**  
83 **effects (DFEs) of ~9,000 single amino acid replacements in TolC under antibiotics,**  
84 **bile salts, colicin E1, or TLS-phage selection. We demonstrated that TolC is highly**  
85 **optimized for the efflux of antibiotics and bile salts. In contrast, under colicin E1**  
86 **and TLS phage selection, TolC sequence is very sensitive to mutations.** This  
87 observation is important in the context of public health where agents such as  
88 bacteriophage and bacterial toxins are favorably viewed as less mutagenic  
89 alternatives to antibiotic therapy.

90 **Results**

91 Creation of a TolC mutant library

92 We measured the evolvability of TolC by quantifying DFEs of all possible single amino  
93 acid replacements in the presence of four physiological stress factors: antibiotics  
94 (piperacillin-tazobactam), bile salts, colicin E1, and TLS phage (Figure 1C-H). First, we  
95 generated a *tolC* deletion strain (*E. coli*- $\Delta$ *tolC*, Methods) which became more sensitive  
96 to both antibiotics and bile salts (Figures S1 and S2) relative to its wild-type parent  
97 strain (BW25113) [22, 23]. The *tolC* deletion strain was also more resistant to both  
98 colicin E1 and TLS phage relative to its wild-type parent (Figure S1). We reintroduced  
99 the *tolC* gene into this strain using a plasmid that has a constitutively active promoter  
100 (pSF-OXB14, Oxford Genetics) and rescued both the antibiotic and bile salt resistance  
101 and the colicin E1 and TLS phage sensitivity of the *E. coli*- $\Delta$ *tolC* strain (Figure S1 and  
102 Figure S2). We mutated all residues except the start codon (471 residues in the mature  
103 TolC protein, and the 21 residue-long signal peptide) of TolC and generated a pool of  
104 ~9,841 (492 sites x 20 aa and a stop codon) mutants (Figure 1C, Figure S3). We cloned  
105 the mutated *tolC* genes into the pSF-OXB14 plasmid and then transformed the *E. coli*-  
106  $\Delta$ *tolC* strain with this pool of plasmids carrying mutated *tolC* genes. We randomly  
107 selected 30 amino acid positions from our library and using Sanger sequencing,  
108 confirmed that all 30 of the mutated sites were randomized and these *tolC* alleles did  
109 not have unintended mutations at other sites (Figure S3). For amplicon sequencing, we  
110 pooled mutants into five sub-libraries and carried out parallel selection and sequencing  
111 experiments (Methods). We deep-sequenced the *tolC* genes in each sub-library by  
112 utilizing the Illumina MiSeq platform and verified that 98.9% of possible amino acid  
113 replacements in the mutant library yielded at least 10 counts when sequenced (Figure  
114 1D-E), with ~1,800 reads per residue or an average of ~90 reads per amino acid  
115 replacement (Figure 1D-E). We also confirmed that frequencies of the mutations in the  
116 *tolC* library did not change significantly when the library was grown in growth media  
117 without selection (Figure 1F,  $p=0.88$ ,  $p<0.001$ , Pearson correlation).

118 Systematic quantification for fitness effects of TolC mutations under positive or negative  
119 selection

120 We measured fitness effects of TolC mutations under selection using a liquid based

121 assay (Figure 1C). In brief, we grew mutant libraries in growth media to saturation,  
122 diluted them to an OD600 of 0.001, and then grew these cultures in the presence of one  
123 of the four selection factors for three hours. Cells were then washed and grown in  
124 nonselective media for six hours. Finally, we harvested plasmids carrying *toIC* mutants  
125 and performed amplicon sequencing to count the surviving *toIC* variants (Methods). The  
126 duration of selection and recovery periods were optimized to maximize the dynamic  
127 range of the measurements and to minimize the chances of losing some alleles during  
128 plasmid harvesting (Methods, Figure S1 and S4). Of note, all concentrations used in  
129 these assays were above the minimum concentrations sufficient to kill wild-type *E. coli*,  
130 except the bile salts. The maximum soluble amount (50 mg/ml) of bile salts in our  
131 selection experiments inhibited growth of the *E. coli*-Δ*toIC* strain but not the wild-type *E.*  
132 *coli* strain (Figure S2). A control experiment with no selection was performed in parallel,  
133 to decouple fitness effects due to growth defects.

134 For calculating fitness, we determined the enrichment of each mutation by comparing  
135 mutation frequencies with and without selection ( $e^i = \log_{10}(\frac{f_{selected}^i}{f_{unselected}^i})$ ), where  $e$   
136 represents the enrichment, and  $f$  the frequency, of mutation  $i$ , Figure 1G). We used the  
137 average fitness of synonymous wild-type mutations as a reference point for defining  
138 relative fitness values ( $s$ ) of each mutation with respect to the wildtype (WT) TolC  
139 sequence ( $s^i = e^i - < e^{WT} >$ ; Figure 1G, green bins). As a control, we compared the  
140 fitness effects of early stop codons (Figure 1G, pink bins) with the *E. coli*-Δ*toIC* strain  
141 supplemented with the *toIC* gene (Figure 1G, green bins) and confirmed that the results  
142 we obtained using our sequencing-based assay matched our observations in batch  
143 culture (Figure S4) both qualitatively and quantitatively. By comparing the enrichments  
144 of mutations in the absence of selection (Methods) relative to the frequencies of  
145 mutations in the library before any growth or selection, we confirmed that TolC  
146 mutations did not have significant fitness effects in the absence of selection (Figure 1F,  
147 H).

148 Figure 2A-B shows the fitness effects for a subset of single amino acid replacements in  
149 TolC in the presence of antibiotics (6 µg/ml), and TLS phage ( $2.5 \times 10^8$  pfu/ml). Figure  
150 S5 summarizes the fitness effects of the entire mutation library under all four selection

151 factors. We found that the fitness effects of mutations in the presence of antibiotics or  
152 bile-salts were mostly neutral (Figure 2A and Figure S5A-B , white pixels) except a  
153 group of mutations increasing sensitivity to antibiotics or bile salts (Figure 2A blue  
154 pixels, Figure 2 C-D insets). Figure 2C and Figure S6 shows the corresponding DFEs.  
155 When we repeated the same assay using 10 times lower dose of antibiotics (0.6  $\mu$ g/ml,  
156 which is still higher than the MIC value of piperacillin-tazobactam for wild-type *E. coli*,  
157 Figure S2A) and bile salts (5 mg/ml), we saw that the DFE in bile salt selection did not  
158 change much but the DFE in antibiotics became slightly narrower further verifying the  
159 robustness of the TolC sequence under antibiotic selection (Figures 2D and S6). None  
160 of the mutations increased resistance to either antibiotics or bile salts suggesting that  
161 the *E. coli* TolC sequence is highly optimized for the efflux function as the TolC  
162 sequence is mostly insensitive to mutations under these conditions, evident from the  
163 corresponding DFEs (Figure 2C, D and Figure S6A, B).

164 On the contrary, in the presence of colicin E1 or TLS phage selection, the majority of the  
165 TolC mutations had large effects on bacterial fitness (Figure 2B, Figure S5C-D, blue  
166 pixels) sometimes making *E. coli* cells more susceptible to colicin E1 or phage induced  
167 death. A subset of mutations increased bacterial resistance to colicin E1 or TLS phage  
168 (Figure 2B, red pixels). Corresponding DFEs were wider relative to the DFEs in  
169 antibiotics and bile salts (Figure 2E-H, Table S1). We repeated these measurements  
170 using ten times higher concentrations of colicin E1 and TLS phage and showed that the  
171 DFEs under these conditions were still wide (Figure 2F,H). These observations  
172 suggested that, under colicin E1 or TLS phage selection (Figure 2 E-H), the *tolC* gene  
173 has the potential to evolve resistance and resides at a sub-optimal fitness state as the  
174 TolC sequence is very sensitive to mutations.

175

176 Relationship between strength of selection and fitness effects

177 We measured fitness effects of TolC mutations using different doses of colicin E1 in  
178 order to measure the relationship between mutational sensitivity and selection strength.  
179 In these experiments, we used increasing concentrations of colicin E1 (0, 5pM, 0.1 nM,  
180 and 2 nM, Figures 3A-E). In addition, we measured fitness effects of TolC mutations in  
181 the presence of TLS phage particles (Figure 3F) and bile salts (Figure S6C). These

182 measurements were done using the Illumina NovaSeq platform and yielded nearly  
183 hundred-fold higher number of reads compared to the MiSeq platform. As the NovaSeq  
184 platform provided large number of sequencing reads, we did not observe any extinct  
185 mutations and we were able to quantify fitness values with greater confidence. We  
186 found that, as the selection strength by colicin E1 increases, the mean values of the  
187 DFEs shift to more negative values and the widths of DFEs become larger (standard  
188 deviation, Figure 3A-E). Similarly, the DFE under phage selection was still wide (Figure  
189 3F) despite the use of 10-fold fewer phage particles compared to our previous  
190 measurements (Figure 2), in agreement with our observations using the MiSeq platform.  
191 On the contrary, the DFEs under bile salt (5mg/ml) selection was narrow, similar to the  
192 DFEs under no selection (Figure S6). Finally under both colicin E1 and phage selection,  
193 we found that a considerable fraction of TolC mutations were resistance-conferring  
194 mutations (Figure 3A-D, F-G, magenta). Almost half of these mutations were early  
195 nonsense substitutions (46% for both colicin E1 selection and TLS phage selection) that  
196 also induced antibiotic or bile salt sensitivity due to disruption of the efflux machinery.  
197 When we excluded stop codon mutations, there were still many (372 mutations  
198 spanning 168 residues for colicin E1 selection, 408 mutations spanning 184 residues for  
199 TLS phage selection) resistance-conferring mutations suggesting that the TolC  
200 sequence was only one mutation away from developing resistance to colicin E1 or TLS  
201 phages (Figure 3G). Using phages or colicin E1 in combination with antibiotics may  
202 potentially reduce the rate of evolution to some extent as early stop codon mutations will  
203 be eliminated by the use of antibiotics. However, many resistance conferring mutations  
204 will still be available and extended use of phages or colicin E1 in clinical settings will  
205 lead to selection of resistant TolC mutants, limiting the success of these therapies in the  
206 long-term.

207 The average fitness effects of TolC mutations under bile salt and antibiotic selection  
208 were both very small and weakly correlated (Figure S6D,  $p= 0.28$  and  $p<0.001$  Pearson  
209 Correlation), making it difficult for us to determine whether efflux efficiencies of these  
210 molecules by TolC were controlled by similar mechanisms. However, the fitness effects  
211 of TolC mutations under phage and colicin E1 selection were broad and significantly  
212 correlated (Figure-4A,  $p= 0.62$  and  $p<0.001$ , Pearson Correlation) suggesting shared

213 infection mechanisms by these agents. As expected, fitness effects under selection by  
214 neither TLS nor colicin E1 showed significant correlation with those measured in the  
215 presence of antibiotics (Figure 4B,C).

216 Structural mapping of mutation-sensitive TolC residues

217 Comparison of the mean fitness effects of TolC missense mutations in the presence of  
218 colicin E1 and TLS phage revealed that resistance to these factors can arise from  
219 mutations in three distinct structural regions (Figure 4D-G), comprising the extracellular  
220 surface of the beta-barrel domain, residues near the so-called “equatorial domain”, and  
221 residues in the periplasmic pore opening [24]. While positional sensitivity to TLS and  
222 colicin E1 selection showed strong overlap, there were several residues that were more  
223 sensitive to one of the two perturbations. Most notably, mutations in a cap of residues  
224 on the extracellular surface of the beta-barrel domain caused resistance to TLS phage,  
225 but were relatively insensitive to colicin E1 (Figure 4D, E). Residues near the equatorial  
226 domain and in the periplasmic pore exhibited sensitivity under both conditions (Figure  
227 4F, G). The effects of mutations under different selection factors may provide some  
228 insight into the mechanisms by which phage and colicin E1 exploit TolC for entry into  
229 the cell.

230 Looking at individual mutations, rather than the mean effects averaged over each  
231 position, provided additional insight. We considered mutations that had fitness effects  
232 larger than three standard deviations from the mean (Methods) and excluded early stop  
233 codon mutations that were equivalent to loss of the *tolC* gene. For this analysis, we also  
234 excluded mutations that did not have consistent fitness effects between experimental  
235 replicates (Methods). We grouped TolC mutations that passed these criteria as  
236 summarized and highlighted mutated residues on the TolC monomer (Figure S7).  
237 Interestingly, many of these residues clustered together suggesting that they might  
238 induce similar changes on the TolC structure due to their physical proximity. We found a  
239 cluster of residues that made *E. coli* cells resistant (Figure S7, dark red residues) to  
240 both colicin E1 and TLS phage when mutated, without significantly changing antibiotic  
241 or bile salt resistance. Strikingly, there was an independent cluster of mutations, that  
242 made *E. coli* cells more sensitive (Figure S7, dark blue residues) to both colicin E1 and

243 TLS, without significantly changing antibiotic or bile salt resistance. One plausible  
244 explanation is that mutations in the helical coiled-coil bundle extending into the  
245 periplasmic space (Figure S7, highlighted in dark red and dark blue) are altering the  
246 interactions of TolC with other proteins involved in colicin E1 or phage entry. Although  
247 not much is known about TLS entry mechanism into bacterial cells, at least three other  
248 proteins are reported to be involved in colicin E1 entry and translocation (TolA, TolQ,  
249 and TolR). Mutations in the residues on the beta barrel (Figure S7, highlighted in dark  
250 red) might be increasing resistance by weakening colicin E1 and TLS binding without  
251 significantly altering antimicrobial efflux. Mutations in the residues highlighted in cyan  
252 (Figure S7) show increased resistance to colicin E1 and TLS phage while making  
253 bacteria more sensitive to both antibiotics and bile salts. This effect can be explained by  
254 misfolding of TolC or a possible constriction of the TolC channel that mechanically or  
255 electrostatically obstructs the passage of all antibiotics, bile salts, colicin E1, and TLS  
256 phage (or viral DNA).

## 257 **Discussion**

258 Utilization of bacterial toxins and bacteriophages to fight bacterial infection has been  
259 proposed for decades and was successfully used in some life-threatening infections. In  
260 fact, bacteriophages that bind the *Pseudomonas aeruginosa* OprM, an efflux protein  
261 homologous to TolC, were recently used to treat a patient with a life threatening pan-  
262 resistant *P. aeruginosa* infection [2]. Our results suggest that if evolutionary aspects are  
263 not taken into account, treatments using phages and bacterial toxins are prone to failure  
264 in the long term, consistent with rapid evolution of phage resistance during treatment of  
265 this patient (Personal communication with Ryland Young, Texas A&M University). Our  
266 study also provides structural clues for understanding mechanisms of TolC mediated  
267 efflux, colicin E1 binding and translocation, and bacterial infection with TLS. Future  
268 structural and functional studies investigating interactions of TolC with its partners and  
269 the spatiotemporal dynamics of these interactions will help to define strategies for  
270 controlling evolutionary outcomes, a key step in addressing problems such as antibiotic  
271 and drug resistance.

## Acknowledgements

We thank Dr. Michael Stiffler for the helpful discussions on making the saturation mutagenesis library. We also thank Dr. William A. Cramer, and Dr. Joe Fralick for providing Colicin E1 plasmid and TLS phage. We thank Dr. Richard Neher and Xiaowei Zhan for their advice on sequence analysis. We thank Ayesha Ahmed for colicin E1 purification and help with sequencing experiments.

## Online Methods

### Growth Media and Strains

*E. coli* cells were grown at 37°C in M9 minimal medium (248510, Difco) supplemented with 0.4% glucose (50-99-7, Fisher Scientific) and 0.2% amicase (82514, Sigma). BW25113 wild-type *E. coli* strain (CGSC#: 7636) and the  $\Delta$ to/C732::kan *E. coli* strain (CGSC#: 11430) were obtained from the Coli Genetic Stock Center. Kanamycin resistance marker was removed from the  $\Delta$ to/C732::kan *E. coli* strain following the protocol in reference [22]. This strain is referred as the  $\Delta$ to/C strain throughout the manuscript. We whole-genome sequenced both the wild-type (BW25113) and the  $\Delta$ to/C *E. coli* strains and confirmed that no other mutations besides the *to/C* deletion were present in the  $\Delta$ to/C strain.

### Saturation mutagenesis assay for the *to/C* gene

pSF-Oxb14 plasmid was obtained from Oxford Genetics (OGS557, Sigma). This plasmid contained a kanamycin resistance cassette and an Oxb14 constitutively open promoter region. The *to/C* gene was PCR amplified from the BW25113 (wild-type) strain using 5'ATTCAAAGGAGGTACCCACCATGAAGAAATTGCTCCCCATTC-3' (forward), and 5'AGAAATCGATTGTATCAGTCTCAGTTACGGAAAGGGTTATGAC-3' (reverse) primers. It was then cloned into the pSF-Oxb14 plasmid using the NEBuilder HiFi DNA Assembly Kit (E5520, New England Biolabs), following the protocol provided by the manufacturer. Bold and underlined nucleotides in primer sequences overlap with the plasmid sequence. The integrated *to/C* gene was confirmed to have no mutations by Sanger sequencing.

Whole gene saturation mutagenesis was performed by two PCR reactions individually for each codon in the *to/C* gene, including the first 22 amino acid long signal sequence. First PCR reaction amplified a portion of the *to/C* gene in the pSF-Oxb14-*to/C* plasmid and randomized the targeted codon with a primer that contained a randomized NNS nucleotide sequence (N stands for A, C, G, or T nucleotides and S stands for G or C nucleotides) for the targeted codon (this PCR product is referred as insert). Second PCR reaction amplified the rest of the pSF-Oxb14-*to/C* plasmid (this PCR product is referred as backbone). Our custom software for designing mutagenesis primers is available at [https://github.com/ytalhatamer/DMS\\_PrimerDesignTool](https://github.com/ytalhatamer/DMS_PrimerDesignTool). Inserts were cloned onto the backbones using the NEBuilder HiFi DNA Assembly Kit (E5520, New England Biolabs and assembled plasmids were transformed into NEB-5-alpha (C2987, New England Biolabs) cells. Plasmid extraction from these cells was done using Nucleospin Plasmid kit (740588, Macharey-Nagel). As this assay produced libraries per each residue, plasmid concentrations were measured and then equimolar amounts of each library were pooled into five sublibraries for 2x250bp paired-end MiSeq

sequencing (residues 2-110, 90-210, 190-310, 290-410, 390-493) and twelve sublibraries for 2x150bp paired-end NovaSeq sequencing (1-40, 36-82, 79-124, 120-166, 162-208, 204-250, 246-292, 288-334, 330-376, 372-418, 414-460, 456-493). Finally, these pooled sublibraries were transformed into  $\Delta$ to/C strain for selection experiments. All growth and selection assays with the library were done using 50  $\mu$ g/ml kanamycin in minimal M9 media.

### **Colicin E1 purification**

A colicin E1 expression vector with IPTG inducible T7 polymerase promoter was kindly provided by Dr. William A. Cramer (Purdue University). Only Colicin E1 was amplified and put back to an empty pET24a plasmid to remove immunity protein. Plasmids were then transformed into BL21-DE3 cells for expression and purification. Cells were grown in TB broth media and colicin E1 was purified first with a size exclusion chromatography. Elutes corresponding to the size of Colicin E1 (~57 kDa) were further purified using a cation exchange chromatography in Sodium borate buffer with a salt gradient of 0-0.3M (NaCl). All fractions are collected and analyzed by SDS-PAGE. Elutes with right band sizes pooled and concentrated using Amicon Centrifugal filters with 30K pore size (UFC803024, Milipore).

### **TLS Phage Harvesting**

TLS phage strain was kindly provided by Dr. Joe Fralick (Texas Tech University). Phage propagation and purification were done following the protocol described in [25]. Briefly, overnight grown bacterial cells were diluted hundred times in 100 mL of LB medium with 5 mM CaCl<sub>2</sub> and incubated 2-3 hours till optical density reached 0.4-0.6. Phage particles were added to the culture and the culture was shaken (at 37°C) until the culture became optically clear. Cell lysates were spun down in 50 mL falcon tubes at 4000 x g and for 20 minutes. Supernatant was filter sterilized using 0.22  $\mu$ m filters. Chloroform was added to the filtered phage solution (10% v/v final chloroform concentration) and the solution was vortexed shortly and incubated at room temperature for 10 minutes. Finally, the phage lysate and chloroform mixture were centrifuged at 4000 x g for 5 minutes. Supernatant was removed, aliquoted, and stored at 4°C.

### **Selection Assay**

We used Piperacillin-Tazobactam (NDC 60505-0688-4, Apotex Corp), bile salts (B8756, Sigma-Aldrich), colicin E1, and TLS phage in selection assays. TolC mutant sublibraries were separately grown overnight in M9 minimal media supplemented with 50  $\mu$ g/mL kanamycin. These cultures were diluted to the optical density of 0.001 in 10 mL of M9 minimal media supplemented with 50  $\mu$ g/mL kanamycin (~5x10<sup>6</sup> cells). Selection agents were added to each sub library and cultures were incubated at 37°C for three hours. All cultures were spun down at 7000 x g for 2 minutes and pellets were resuspended in fresh M9 minimal media supplemented with 50  $\mu$ g/mL kanamycin. These cultures were then incubated at 37°C with shaking for six hours. Following this step, libraries were centrifuged at 7000 x g for 2 minutes and pellets were collected for plasmid purification. Different regions of the to/C genes were amplified with PCR and indexed using Illumina Index sequences (Supplementary Data). These regions were spanning residues 2-110, 90-210, 190-310, 290-410, and 390-493 for 2x250bp paired-

end MiSeq sequencing. For 2x150 bp paired-end NovaSeq sequencing, we amplified and indexed the residues 2-40, 36-82, 79-124, 120-166, 162-208, 204-250, 246-292, 288-334, 330-376, 372-418, 414-460, and 456-493.

### Sequence analysis

Paired ended sequencing reads were first merged using the FLASH tool [26] (Customized parameters: -m 40 -M 100). Reads covering primers overlapping with the upstream and downstream of the amplified regions of *to/C* were excluded. Sequence reads were compared to the wild-type *to/C* sequence and mutations were listed. Sequence reads that had mutations in more than one residue were excluded from the analysis. Synonymous mutations yielding the same amino acid replacement were grouped together. Frequency of each mutation was calculated by dividing number of counts of for that mutation with number of all reads, including alleles with multiple mutations ( $f^i = \frac{N^i}{N_{all}}$ ). For calculating fitness, we first determined enrichment of each mutation by comparing mutation frequencies with and without selection ( $e^i = \log_{10}(\frac{f^i_{selected}}{f^i_{unselected}})$ ); e stands for enrichment and f stands for frequency, Figure 1G). Since randomized mutations of each residue created traceable mutations synonymous to the wild-type protein sequence, we were able to use average fitness of synonymous wild-type mutations for defining relative fitness values of each mutation with respect to the wild-type (WT) TolC sequence ( $s^i = e^i - < e^{WT} >$ ; s stands for fitness, Figure 1K, green bins). As a sanity check, we compared the fitness effects of early stop codons with the phenotype of the *E. coli*- $\Delta to/C$  strain (Figure 1G, pink bins) and confirmed that the results we obtained our sequencing-based assay matched our observations in batch culture (Figure S4) both qualitatively and quantitatively. By comparing the enrichments of mutations in the absence of selection relative to the frequencies of mutations in the library before any growth or selection, we confirmed that TolC mutations did not have significant fitness effects in the absence of selection (Figure 1F, H). Our source code for data analysis is available at

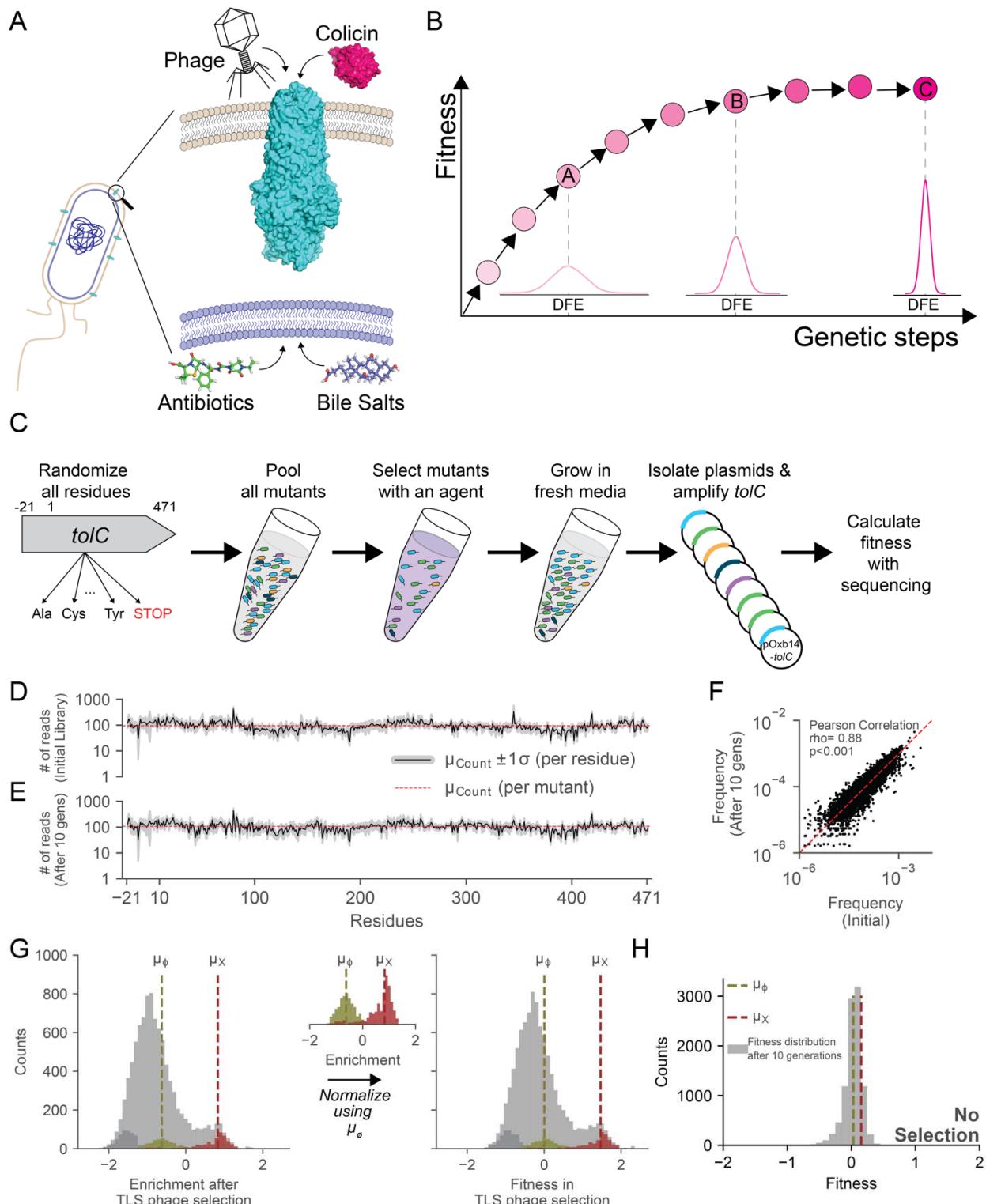
[https://github.com/ytalhatamer/DMS\\_DataAnalysis](https://github.com/ytalhatamer/DMS_DataAnalysis).

## References

1. Petrovic Fabijan, A., et al., *Safety of bacteriophage therapy in severe *Staphylococcus aureus* infection*. Nat Microbiol, 2020. **5**(3): p. 465-472.
2. Chan, B.K., et al., *Phage treatment of an aortic graft infected with *Pseudomonas aeruginosa**. Evol Med Public Health, 2018. **2018**(1): p. 60-66.
3. Schooley, R.T., et al., *Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant *Acinetobacter baumannii* Infection*. Antimicrob Agents Chemother, 2017. **61**(10).
4. Pagie, L. and P. Hogeweg, *Colicin diversity: a result of eco-evolutionary dynamics*. J Theor Biol, 1999. **196**(2): p. 251-61.
5. Cascales, E., et al., *Colicin biology*. Microbiol Mol Biol Rev, 2007. **71**(1): p. 158-229.
6. Andersen, C., C. Hughes, and V. Koronakis, *Chunnel vision. Export and efflux through bacterial channel-tunnels*. EMBO Rep, 2000. **1**(4): p. 313-8.
7. Paulsen, I.T., et al., *A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria*. FEMS Microbiol Lett, 1997. **156**(1): p. 1-8.
8. Johnson, J.M. and G.M. Church, *Alignment and structure prediction of divergent protein families: periplasmic and outer membrane proteins of bacterial efflux pumps*. J Mol Biol, 1999. **287**(3): p. 695-715.
9. Koronakis, V., J. Eswaran, and C. Hughes, *Structure and function of TolC: the bacterial exit duct for proteins and drugs*. Annu Rev Biochem, 2004. **73**: p. 467-89.
10. Piddock, L.J., *Multidrug-resistance efflux pumps - not just for resistance*. Nat Rev Microbiol, 2006. **4**(8): p. 629-36.
11. Ayhan, D.H., et al., *Sequence-Specific Targeting of Bacterial Resistance Genes Increases Antibiotic Efficacy*. PLoS Biol, 2016. **14**(9): p. e1002552.
12. Cremers, C.M., et al., *Bile salts act as effective protein-unfolding agents and instigators of disulfide stress in vivo*. Proc Natl Acad Sci U S A, 2014. **111**(16): p. E1610-9.
13. Hamner, S., et al., *Bile salts affect expression of *Escherichia coli* O157:H7 genes for virulence and iron acquisition, and promote growth under iron limiting conditions*. PLoS One, 2013. **8**(9): p. e74647.
14. Bloom, J.D., et al., *Evolution favors protein mutational robustness in sufficiently large populations*. BMC Biol, 2007. **5**: p. 29.
15. Wagner, A., *Distributed robustness versus redundancy as causes of mutational robustness*. Bioessays, 2005. **27**(2): p. 176-88.
16. Stiffler, M.A., D.R. Hekstra, and R. Ranganathan, *Evolvability as a function of purifying selection in TEM-1 beta-lactamase*. Cell, 2015. **160**(5): p. 882-892.
17. Fisher, R.A.S., *The genetical theory of natural selection*. 1930, Oxford: Clarendon Press.
18. Losos, J.B., et al., *Evolutionary biology for the 21st century*. PLoS Biol, 2013. **11**(1): p. e1001466.
19. Fisher, R., *The Genetical Theory of Natural Selection*. Oxford Clarendon Press, 1930.
20. Darwin, C., *On the Origin of Species*. 1859.

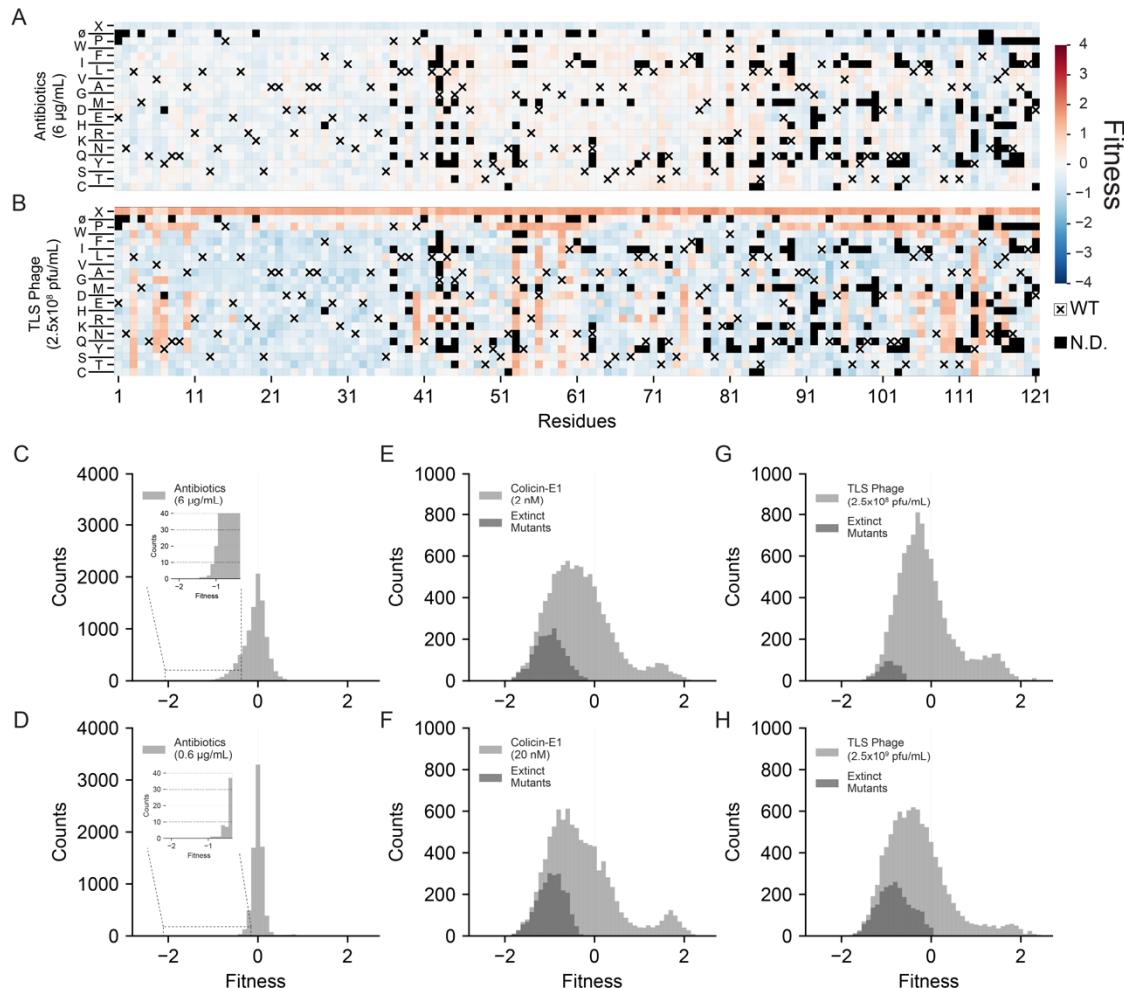
21. Chevereau, G., et al., *Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance*. PLoS Biol, 2015. **13**(11): p. e1002299.
22. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products*. Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.
23. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection*. Mol Syst Biol, 2006. **2**: p. 2006 0008.
24. Koronakis, V., et al., *Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export*. Nature, 2000. **405**(6789): p. 914-9.
25. Bonilla, N., et al., *Phage on tap-a quick and efficient protocol for the preparation of bacteriophage laboratory stocks*. PeerJ, 2016. **4**: p. e2261.
26. Magoc, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies*. Bioinformatics, 2011. **27**(21): p. 2957-63.

## Figures and Figure Legends



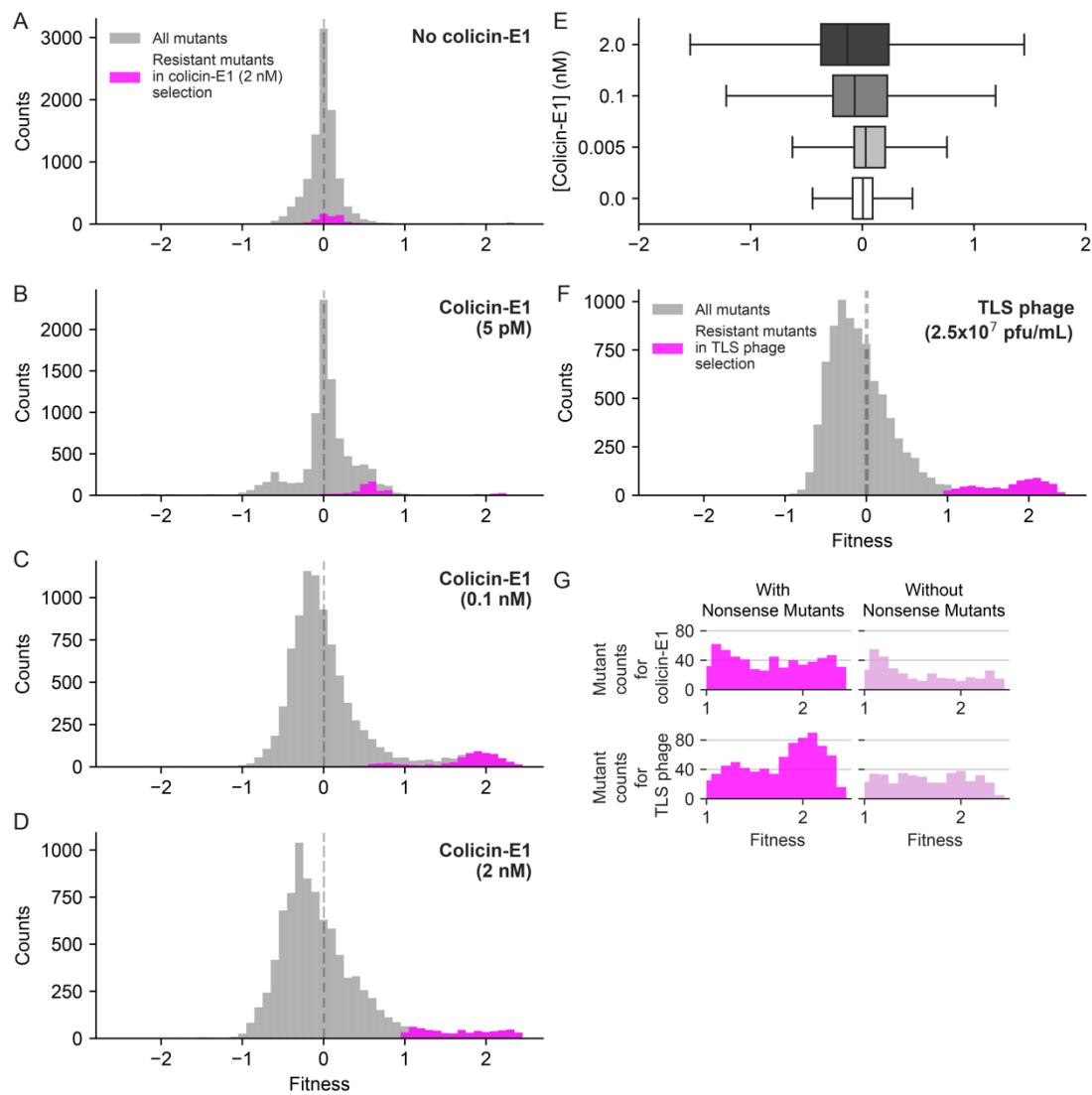
**Figure-1. Creation of TolC mutant library (A)** TolC is an outer membrane protein in *E. coli* that is involved in the efflux of antibiotics and bile salts. TolC protein is also exploited

as a receptor by TLS phage and colicin-E1, a bacterial toxin. **(B)** Fisher's Geometric theorem predicts evolvability of biological systems using the widths of distributions of fitness effects (DFE). Large DFEs are indicators of high evolvability. Narrow DFEs are observed when biological systems are robust to genetic perturbations and the narrow width suggests lower potential to evolve. **(C)** Experimental procedure for whole gene saturation mutagenesis and fitness measurements. Deep Mutational Scanning of TolC was done by randomization of all 493 residues (including the signal sequence between residues -22 to -1, except the start codon) to 19 other amino acids and the stop codon. All mutants were pooled together and grown under selection of one of the four agents (antibiotics, bile salts, colicin-E1 or TLS phage). The *to/C* alleles were then harvested after a brief recovery growth in plain growth media, and frequency of mutations were calculated by deep sequencing of the *to/C* alleles. **(D-E)** Number of reads for all TolC mutations are plotted before (D) and after (E) growing the TolC mutant library for 10 generations (lower panel) in minimal media. Black line represents the mean ( $\mu$ ) number of reads per mutation at each residue. Gray area around the black line shows  $\pm 1$  standard deviation from the mean. Horizontal dashed red line marks the mean number of reads for all TolC mutations. **(F)** Frequencies of all *to/C* mutations before and after 10 generations of growth in minimal media were highly correlated (Pearson, rho= 0.88 and p< 0.001). Red diagonal dashed line shows  $y=x$  line. **(G)** We calculated enrichment of each mutation by comparing mutation frequencies with and without selection ( $e^i = \log_{10}(\frac{f_{selected}^i}{f_{unselected}^i})$ ;  $e$  stands for enrichment and  $f$  stands for frequency). Light green histogram represents enrichment values of mutations synonymous to the wild-type TolC protein sequence. Light red histogram represents mutants with early stop codons mutations. Mean values for the stop codon mutations ( $\mu_x$ ) are represented with the vertical dashed red lines. Dark grey colored histograms on the left shows the mutants that went extinct after the selection. Their fitness values were calculated by assuming that their counts were equal to 0.01 after selection in order to manually separate them from the rest of the mutations. We used average fitness of synonymous wild type mutations ( $\mu_s$ ; dashed vertical green line) for defining relative fitness values of each mutation with respect to the wild type (WT) TolC sequence ( $s^i = e^i - < e^{WT} >$ ;  $s$  stands for fitness). **(H)** Distribution of fitness effects for the TolC mutant library after growth in minimal media (~10 generations) without any selection.



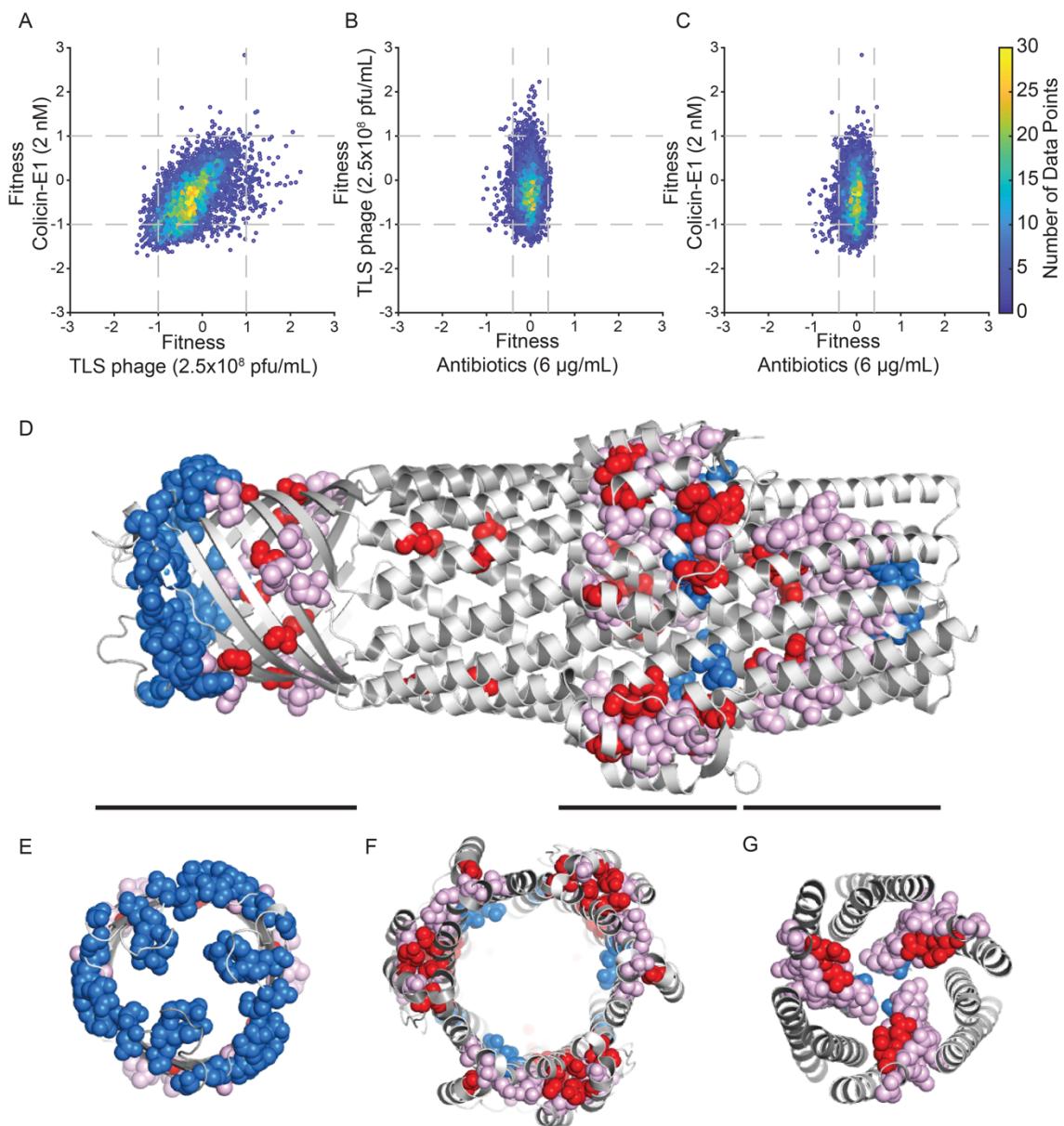
**Figure-2. Mutant library selection using antibiotics and TLS Phage** Heatmaps summarizing a subset of fitness values of TolC mutations under selection of (A) Antibiotics (6 µg/mL), (B) TLS Phage (2.5x10<sup>8</sup> pfu/mL). Columns within each matrix represent the TolC residues (from residue 1 to 121), and columns represent mutations to other codons causing synonymous(Ø), nonsynonymous, or nonsense (X) mutations. Protein sequence of the wild-type TolC are represented by white pixels with cross on them. Mutants with low read counts (threshold=  $\mu-1.5\bar{x}$ ) in the initial TolC library or after growth in plain media (untreated) were excluded from fitness calculations and represented by black pixels. All other calculated fitness values were colored from dark blue (increased susceptibility) to white (neutral effects) to dark red (increased resistance). Top row of each heatmap shows the effect of early stop codon mutations (X). Second rows from the top show the effect of silent mutations synonymous to native codon in TolC. Under antibiotics selection (A), nonsense mutations increased susceptibility, whereas under TLS phage selection (B), nonsense mutations increase resistance. **(C-H)** Distribution of fitness effects (DFEs) for different selection agents. For every selection agent, DFEs were calculated under two different selection strengths. DFEs under antibiotic selection are narrow and centered around neutrality ( $s = 0$ ) regardless of the selection strength, with tails extending to the left (increased sensitivity, insets). Under both colicin-E1 and TLS phage selections, DFEs were wide and mean

fitness effects of mutations were negative, suggesting that TolC was not robust to mutations under selection to these agents. Under both colicin-E1 and phage selection, many mutations were initially present in the TolC library but went extinct after the selection (dark gray bins). Although it is not possible to calculate fitness in these cases, in order to show them on the histograms, we set their final counts to 1 (pseudocount) and calculated a fitness value such that they are still visible. Under both colicin-E1 and phage selections, DFEs were bimodal with a second peak corresponding to mutations increasing resistance. Means and standard deviations of light gray histograms are tabulated in Table S1.



**Figure 3. Impact of selection strength on DFE (A-D)** Selection strength alters the shape and (E) width of the distribution of fitness effects under colicin-E1 selection. We measured fitness effects of TolC mutations by Illumina NovaSeq sequencing platform which yielded ~100 times more reads per mutation, compared to Illumina MiSeq platform, and increased resolution of our fitness measurements. DFEs for TolC mutations under selection with (A-D) increasing concentrations of colicin-E1, and (F) TLS Phage (2.5x10<sup>7</sup> PFU/mL). Magenta colored bins in panels A to D highlight resistance-conferring mutations that had fitness values larger than 1 (10-fold change in frequency) under selection with 2nM of colicin-E1. Magenta colored bins in panel F highlight resistance-conferring mutations that had fitness values larger than 1 under phage selection. (G) (left) Histograms of all resistance-conferring mutations under colicin-E1 (2nM, 685 mutations) selection and TLS phage selection (761 mutations). (right) Histograms of all resistance-conferring mutations, excluding stop codon

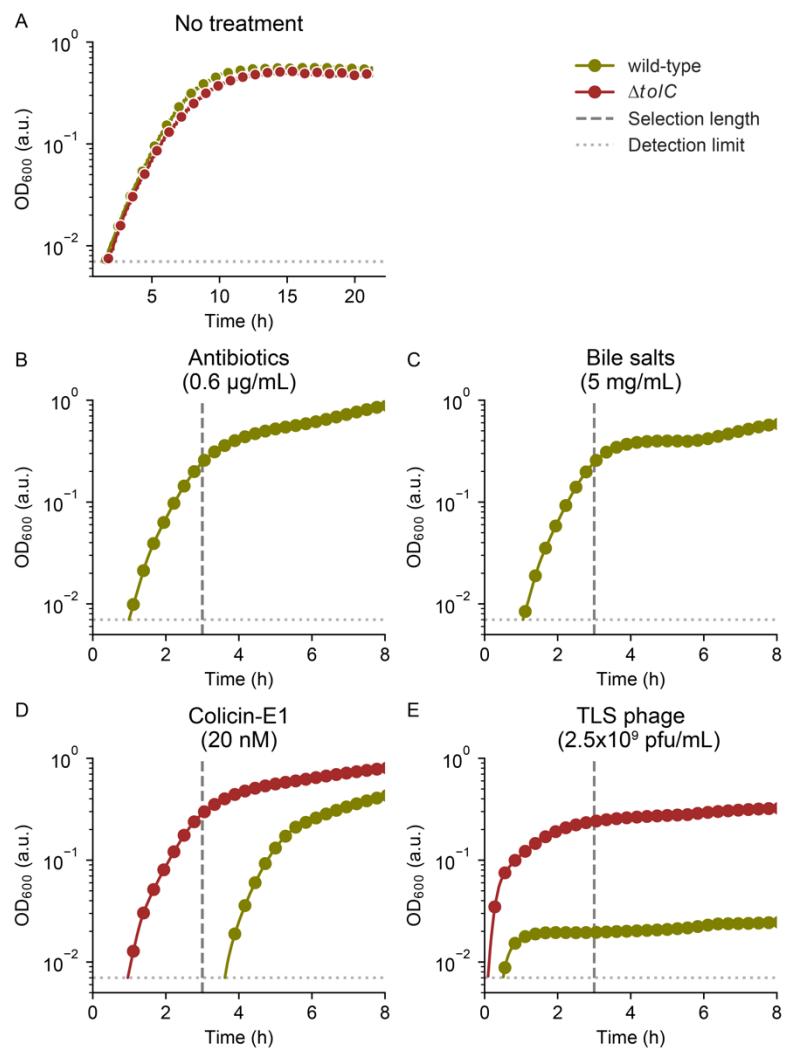
mutations, under colicin-E1 (2nM, 372 mutations) selection and TLS phage selection (408 mutations).



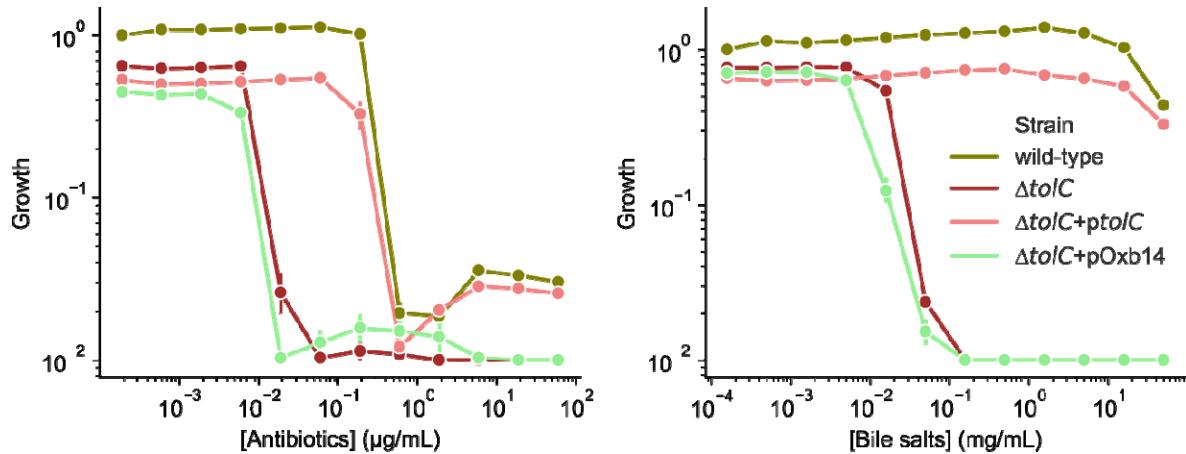
**Figure 4. Analysis of mutations from selection experiments (A-C)** Comparison of fitness effects of TolC mutations under different selections. These fitness values correspond to consistent values used to plot histograms in Figure 2C-H. Horizontal and vertical dashed lines indicate the significance thresholds ( $\pm 1.5$  standard deviations) for each selection condition, which correspond to  $\sim 10$  fold increase or decrease in frequency relative to wild-type TolC (fitness values  $>+1$  and  $<-1$ ) under colicin-E1 or TLS phage selection. Significance threshold under antibiotic selection were  $\sim 2.5$  fold increase or decrease in frequency relative to wild-type TolC (fitness values  $> +0.4$  and  $< -0.4$ ). **(D)** Side-view of the TolC trimer. The 47 most sensitive residues (representing the top 10%) for TLS and colicin E1 are shown in spheres. Positions that show sensitivity to TLS only are colored blue, sensitivity to colicin E1 only in red, and sensitivity to both perturbations in pink. Black bars correspond to slices along the pore

axis, shown in (E-G) highlighting the three structural regions where sensitive residues are located.

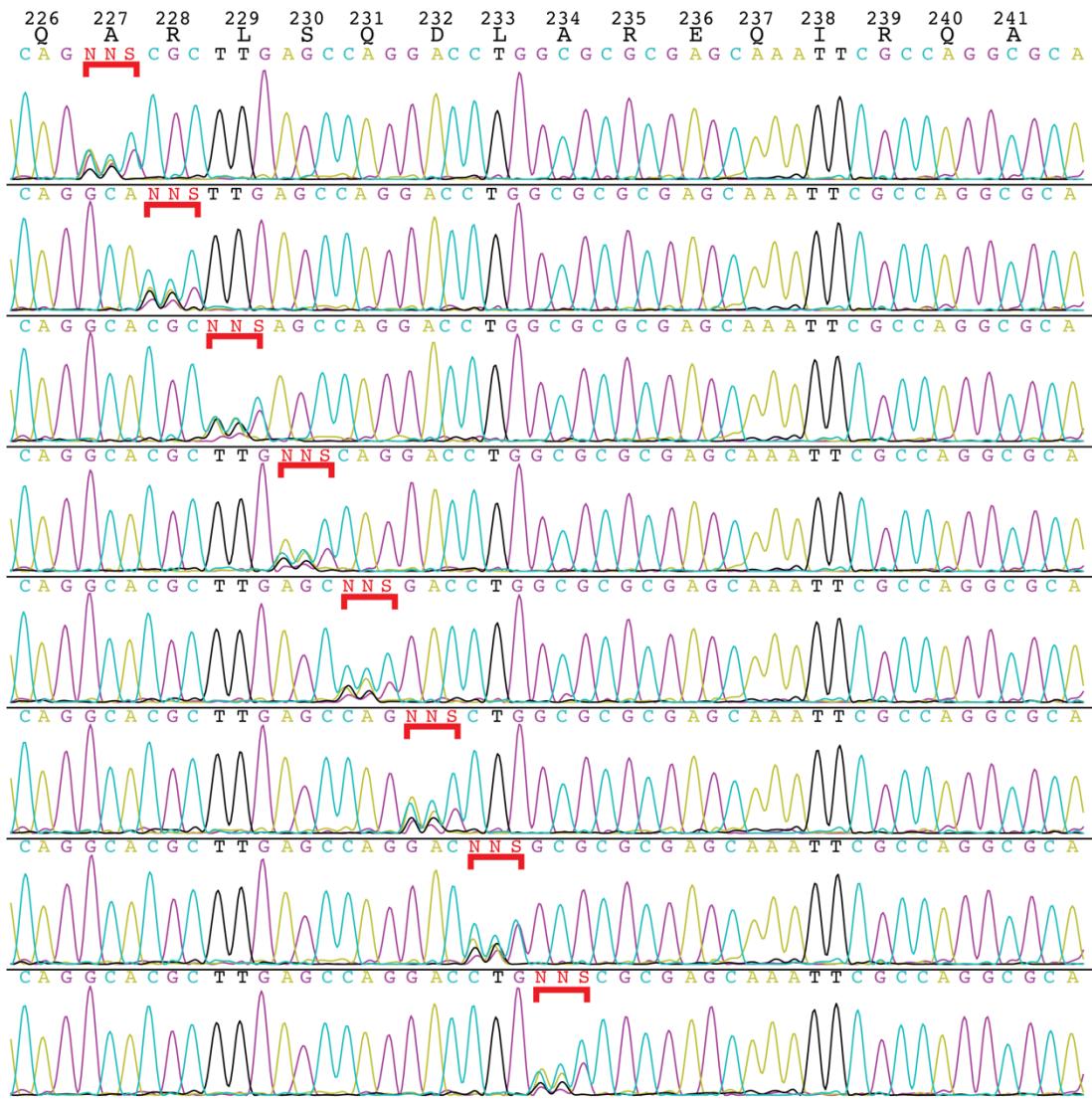
## Supplementary Figures



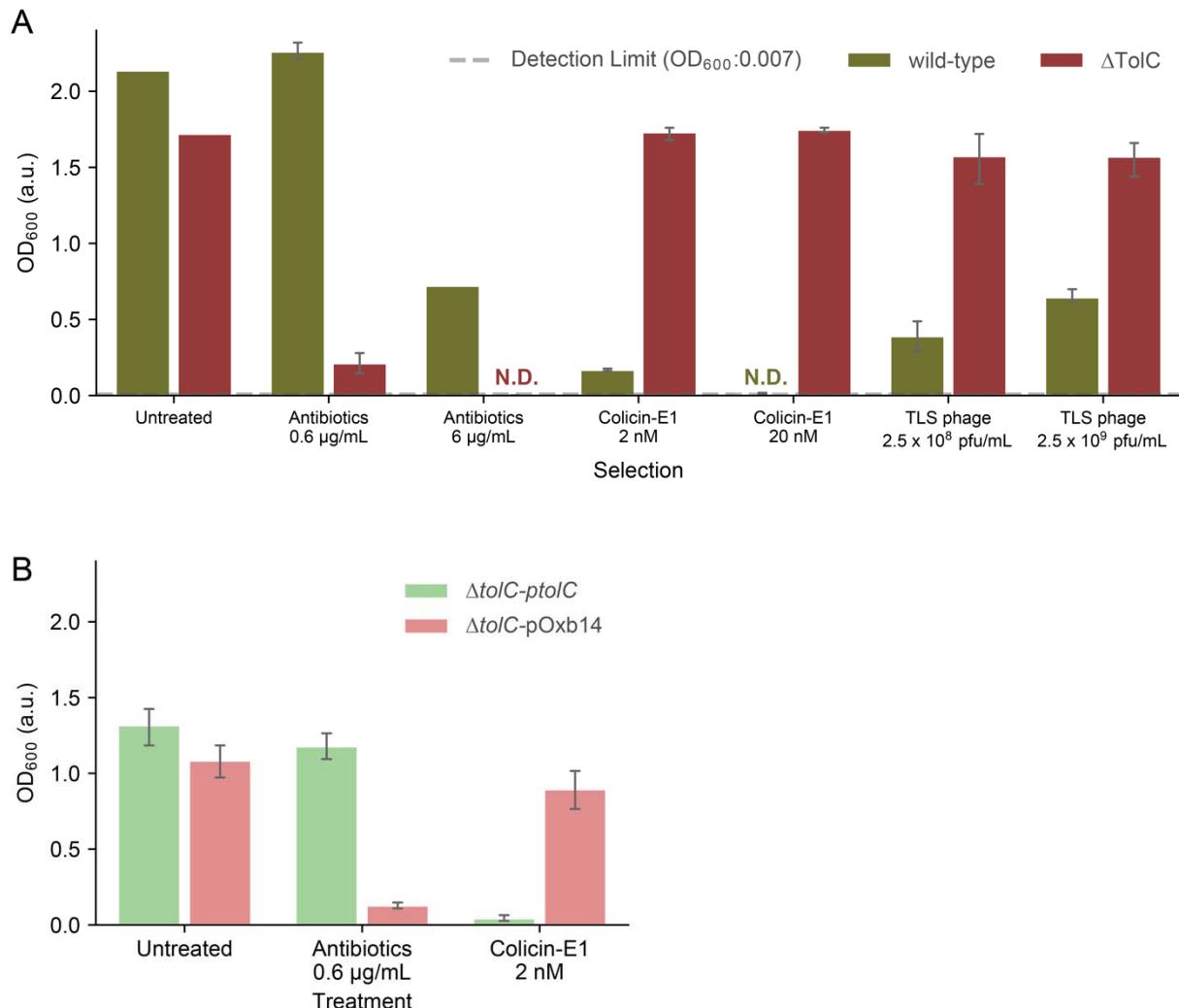
**Figure S1. Deletion of the *tolC* gene is costly under antibiotic and bile salt selection and beneficial under colicin-E1 and TLS Phage selection.** (A) Growth curves of wild-type and  $\Delta tolC$  strains without selection pressure. Dark green colored lines represent growth curves of the wild type (BW25113) *E. coli* strain. Dark red colored lines represent growth curves of the BW25113 *E. coli* strain with *tolC* gene deletion ( $\Delta tolC$ ). In our fitness assays, we used a duration of three hours for selection (vertical grey dashed line) in order to maximize the fitness difference between the wild type *E. coli* and *E. coli*: $\Delta tolC$ . Horizontal gray dotted line represents the detection limit of the spectrophotometer used (OD<sub>600</sub>: 0.007).



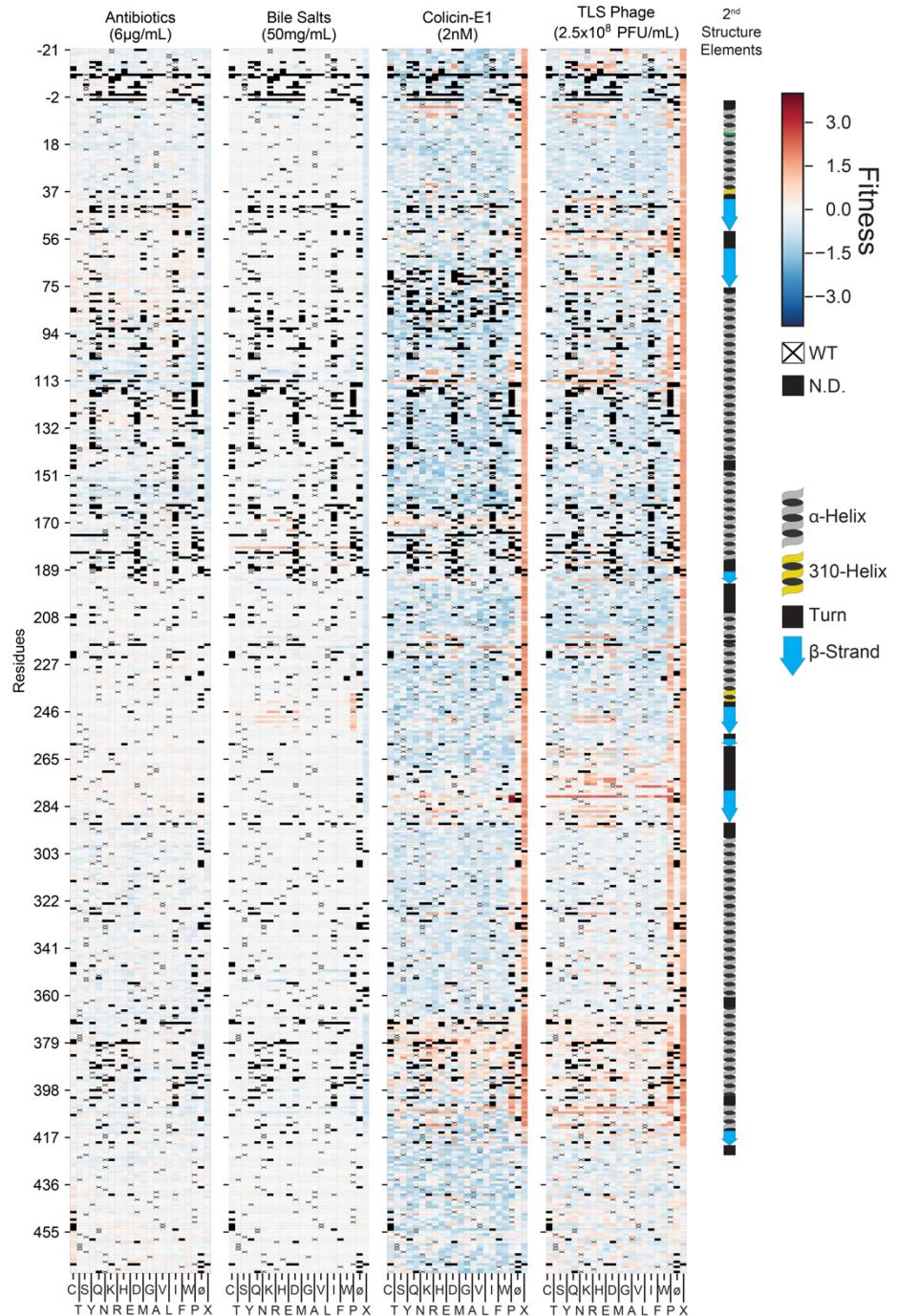
**Figure S2. Supplementing the *tolC* gene into the *E. coli*- $\Delta\text{tolC}$  rescues antibiotic and bile salt resistance.** Dose response curves are reported under antibiotics (left panel) and bile salts(right panel) selection for wild-type BW25113,  $\Delta\text{tolC}$ ,  $\Delta\text{tolC}+\text{pOxb14}$  and  $\Delta\text{tolC}+\text{ptolC}$  strains. Growth of the strains are measured at 600nm every 30 minutes. Area under the growth curves are calculated for each antibiotic and bile salts concentration. Growth values are reported on y-axis after growth values are normalized using wild-type growth in the absence of selection.



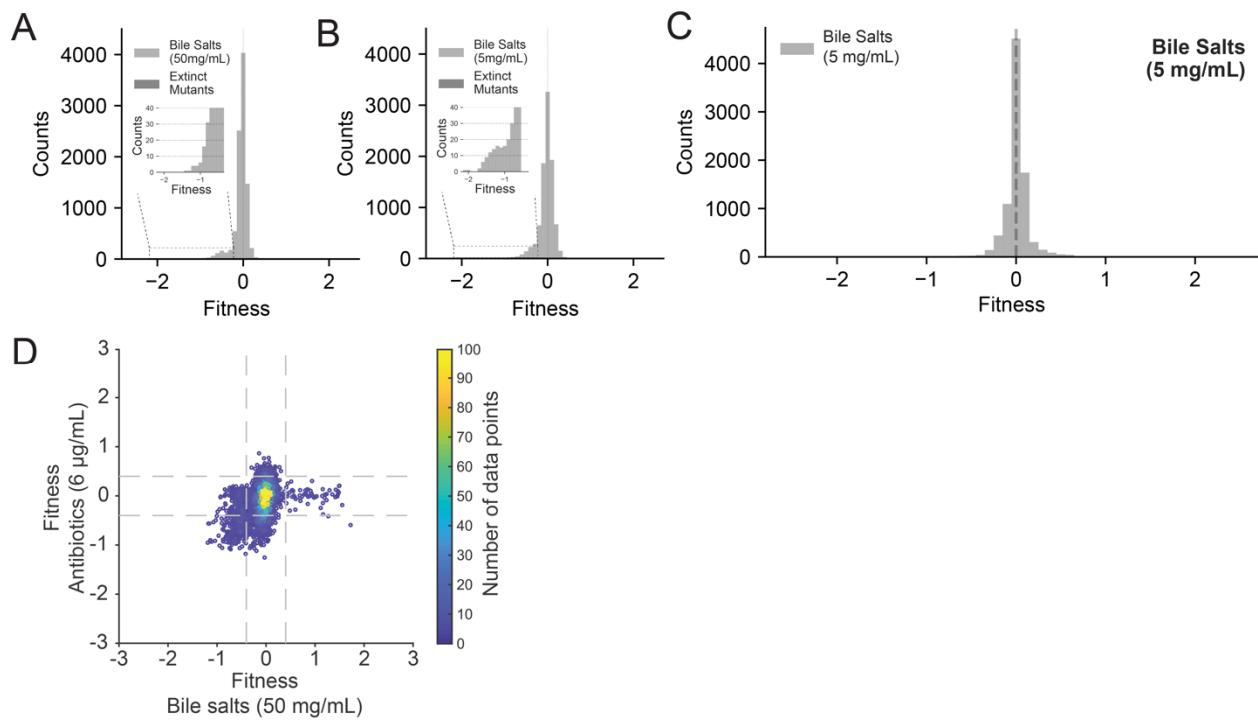
**Figure S3. Confirmation of saturation gene mutagenesis with Sanger sequencing.** Native amino acids for mutated positions (residues 227-234) are shown in black in the top row. Randomization of the intended residue was done separately for each position. Chromatograms demonstrate successful randomization of each codon to NNS (Any nucleotide in first two position and either G or C in the third position). In chromatograms, Cyan color was used for Cytosine, Black for Thymidine, Magenta for Guanine, and Yellow for Adenine. Sequence of each fragment is shown above the chromatogram. Randomized regions are highlighted using a red squared bracket.



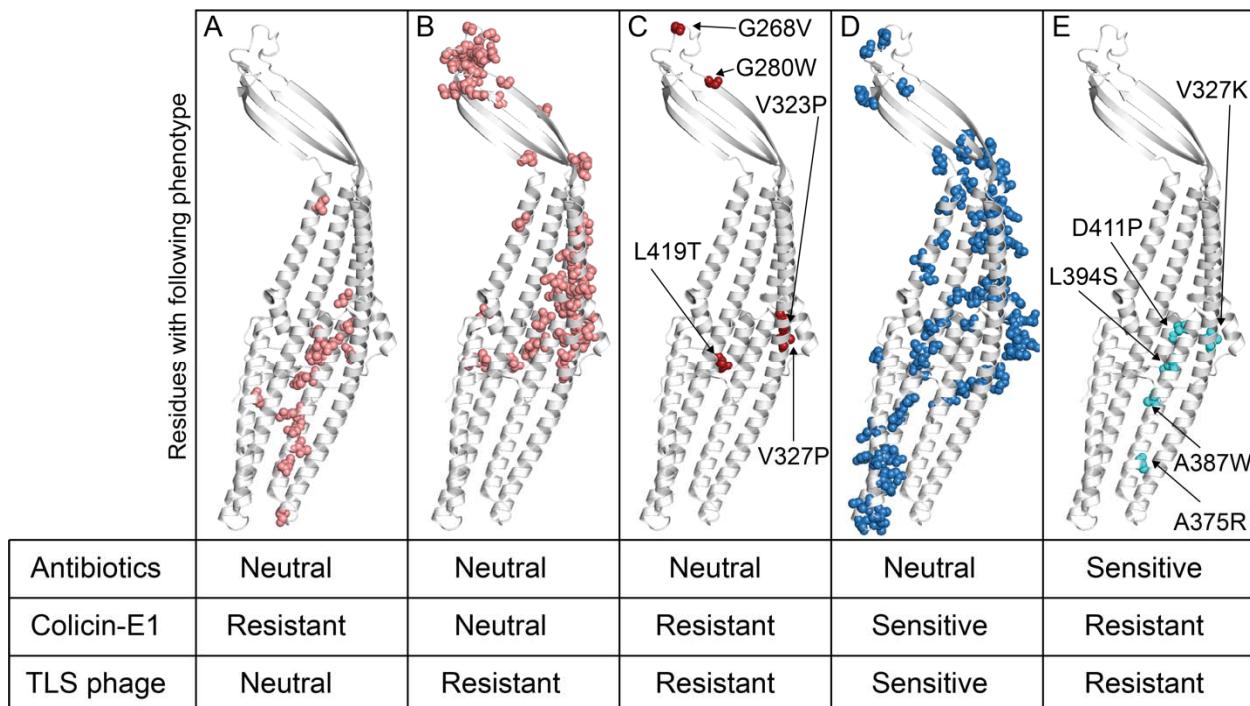
**Figure S4. Effects of antibiotic, colicin-E1 and TLS phage selections are quantified with an optical density based assay for wild-type *E. coli* (green) and *E. coli*:ΔtolC (red).** We plotted cell densities with and without selection. **(A)** Growth values at two concentrations of antibiotics, colicin-E1 and TLS phage were tested. No growth has been observed for ΔtolC strain in Antibiotics (6 µg/mL) selection and for wild-type strain for colicin-E1 (20nM) selection which are represented as not determined (N.D. with respective colors). Detection limit for spectrophotometer plotted as horizontal gray dotted line. **(B)** Growth is measured in the presence of antibiotics (piperacillin-tazobactam) and colicin-E1 for the *E. coli*:ΔtolC strain supplemented with an empty plasmid (red) and with a plasmid carrying the wild type *tolC* gene (green).



**Figure S5. Fitness effects under four different stress conditions are plotted as heatmaps.** Y axis in the heatmaps show the residues. Columns in heatmaps represents synonymous(ø), nonsynonymous and stop codon (X) mutations on each residue. Known secondary structure elements of TolC protein are highlighted in color.



**Figure S6: Distribution of fitness effects (DFEs) for bile salts selection. (A-B)**  
DFEs were calculated under two different selection strengths. DFEs under bile salts selection are narrow and centered around neutrality ( $s = 0$ ) regardless of the selection strength, with tails extending to the left (increased sensitivity, insets). (C) DFE of bile salts selection measured in Illumina NovaSeq platform. Distribution is narrow and centered around neutrality as well. (D) Comparison of fitness effects of TolC mutations under antibiotics (6 µg/mL) and bile salts (50 mg/mL) selections. There is a weak correlation in fitness values under these selection conditions ( $\rho = 0.28$  and  $p < 0.001$  Pearson Correlation). Vertical and horizontal dashed lines represent three standard deviation from mean ( $\pm 0.4$ ).



**Figure S7: Mutated residues that caused significant fitness changes under at least one of the three selection factors are highlighted in color on monomeric TolC structure.** (Table S2) **(A)** Mutations in pink colored residues confer resistance to colicin-E1 without causing significant fitness changes under other selection conditions. **(B)** Mutations in pink colored residues confer resistance to TLS phage without causing significant fitness changes under other selection conditions. **(C)** Mutations in red colored residues increase resistance to both colicin-E1 and TLS phage without disrupting efflux of antibiotics. **(D)** Mutations in blue colored residues increase sensitivity to both colicin-E1 and TLS phage without disrupting efflux of antibiotics. **(E)** Mutations in cyan colored residues increase resistance to both colicin-E1 and TLS phage and disrupt efflux of antibiotics. These mutations likely cause misfolding of TolC or blockage of the TolC channel as their fitness effects are reminiscent of the loss of the *tolC* gene.

**Table S1:** Mean and standard deviation values for DFEs under all selection conditions: antibiotics (0.6 and 6  $\mu\text{g/mL}$ ), bile salts (5 and 50  $\text{mg/mL}$ ), colicin-E1 (2 and 20nM), and TLS phage ( $2.5 \times 10^8$  and  $2.5 \times 10^9$  pfu/mL).

**Table S2:** Fitness values of mutations represented in Figure S7 are tabulated. There are six sheets in the excel file. First sheet shows the consistent data that has similar fitness values in two different experiments. Remaining five sheets show fitness values for the mutations represented in Figure S7A-E.