The RND efflux pump EefABC is highly conserved within lineages of *E. coli* commonly associated with infection.

Running Title: EefABC; the seventh RND pump in *E. coli*

Hannah L. Pugh[1,2+], Elizabeth M. Darby[1+], Leah Burgess[1], Abigail L. Colclough[1], Asti-Rochelle Meosa John[1], Steven Dunn[1], Christopher Connor[1,3], Eoughin A. Perry[1], Alan McNally[1], Vassiliy N. Bavro[4] and Jessica M. A. Blair*[1]

+ Joint first author

*Corresponding Author

[1] Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT

[2] The Francis Crick Institute, 1 Midland Road, London, NW1 1AT

[3] Department of Microbiology and Immunology at the Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria, 3010, Australia

[4] School of Life Sciences, University of Essex, Colchester, CO4 3SQ, UK

Abstract

Tripartite resistance-nodulation-division (RND) efflux pumps confer multidrug resistance (MDR) in Gram-negative bacteria and are critical for many physiological functions including virulence and biofilm formation. The common laboratory strain of *E. coli,* K-12 MG1655 has six recognised RND transporters participating in tripartite pump formation (AcrB, AcrD, AcrF, CusA, MdtBC, and MdtF). However, by studying >20,000 *E. coli* genomes we show that *E. coli* belonging to phylogroups B2, D, E, F and G, which are commonly associated with infection, possess an additional, seventh RND transporter, EefB. It is found in a five gene operon, *eefRABCD,* which also encodes a TetR family transcription factor, a periplasmic adapter protein, an outer membrane factor and major facilitator superfamily pump. In contrast, *E. coli* from phylogroups A, B1 and C, generally containing environmental and commensal strains, do not encode the operon and instead encode an uncharacterised ORF, *ycjD*. In phylogroups where the *eefRABCD* operon is present it was very highly conserved. In fact, conservation levels were comparable to that of the major *E. coli* RND efflux

system AcrAB-TolC, suggesting a critical biological function. Protein modelling shows that this pump is highly divergent from endogenous *E. coli* RND systems with unique structural features, while showing similarities to efflux systems found in *Pseudomonas aeruginosa*. However, unlike other major RND efflux systems, EefABC does not appear to transport antimicrobials and instead may be important for infection or survival in the host environment.

Importance

Efflux pumps are molecular machines that export molecules out of bacterial cells. The efflux pumps belonging to the RND family are particularly important as they export antibiotics out of Gram-negative bacterial cells, contributing to antibiotic resistance. The important human pathogen, *E. coli*, has been previously reported to have six RND pumps. However, we show that phylogroups of *E. coli* commonly associated with infection encode a seventh RND pump, EefABC which is highly conserved, suggesting an important biological function. While the function of EefABC in *E. coli* remains to be resolved, it does not seem to transport antimicrobial compounds. These findings are important because they reveal a new RND pump, potentially involved in virulence and survival in the host, that could represent a new therapeutic target. Additionally, it again shows that laboratory type strains of common bacterial pathogens are not representative of those that are infection causing.

Introduction

Escherichia coli is a leading cause of invasive bacterial infections in humans, causing a range of diseases from urinary tract infections to haemorrhagic shock (Holmes et al., 2021). However, *E. coli* is also found in the wider environment and is a common commensal, colonising the gastrointestinal tract of both humans and animals. *E. coli* is a genetically diverse species that is divided into phylogroups (A, B1, B2, C, D, E, F and G) which are determined by genetic similarity (Denamur et al., 2021). Though virulence is not limited to specific phylogroups (Clermont et al., 2000, Gordon and Cowling, 2003), extra-intestinal infection is most commonly associated with phylogroups B2 and D (Denamur et al., 2021), while commensal or environmental lifestyles are mainly associated with phylogroups A and B1 (Tenaillon et al., 2010). In addition to phylogroups, *E. coli* isolates are also classified by sequence types (ST),

69 determined by multi-locus sequence typing (MLST) (Maiden et al., 1998). Pandemic
70 clonal STs such as ST131 are responsible for high incidence rates of extra-intestinal
71 *E. coli* infections and the carriage of genes conferring multidrug resistance (Shaik et
72 al., 2017, Li et al., 2023).
73 Gram-negative bacteria, such as *E. coli*, possess efflux pumps of the
74 resistance-nodulation-division (RND) family which are best known for the export of
75 antimicrobials and biocides (Chen et al., 2003, Sanchez et al., 1997, Nishino and
76 Yamaguchi, 2001, Alav et al., 2021) with overexpression of these pumps conferring
77 multidrug resistance (MDR) (Swick et al., 2011, Chetri et al., 2019). RND pumps form
78 tripartite complexes with periplasmic adaptor proteins (PAP) and outer membrane
79 factors (OMF) to span the inner membrane, periplasmic space, and the outer
80 membrane (Alav et al., 2021). In addition to MDR, RND pumps have been implicated
81 in a wide range of additional functions, including virulence (Nishino et al., 2006, Blair
82 et al., 2015, Padilla et al., 2010), biofilm production (Baugh et al., 2014), decreased
83 susceptibility to dyes, bile salts and fatty acids (Nishino and Yamaguchi, 2001, Mateus
84 et al., 2021), export of polyamines and quorum sensing molecules (Chan et al., 2007),
85 export of bacterial metabolites (Wang-Kan et al., 2021), copper and ion homeostasis
86 (Grass and Rensing, 2001, Franke et al., 2001), and motility (Bazzini et al., 2011, Yang
87 et al., 2011).
88 Many Gram-negative bacteria encode multiple RND efflux pumps, with differing
89 substrate specificities which are often expressed only under specific environmental
90 conditions. However, the number of RND pumps is highly variable between bacterial
91 lineages. The human-restricted pathogen *Neisseria gonorrhoeae* possesses only a
92 single RND pump (MtrCDE), while bacterial species that can colonise multiple habitats
93 generally possess more. *Salmonella enterica* serovar Typhimurium has five RND
94 pumps, whereas *E. coli* is generally reported to have six: AcrB, AcrD, AcrF, MdtBC,
95 MdtF and CusA (Nishino and Yamaguchi, 2001). However, the number of RND pumps
96 can also vary within a genus. For example, we have recently shown that species
97 across the *Acinetobacter* genus possess between two and nine RND pumps, with
98 species most-commonly associated with human infection tending to encode more
99 RND efflux pumps (Darby et al., 2023). A study by Ma *et al.* found loss of MtrC function
100 correlated with increased drug susceptibility in *N. gonorrhoeae* isolated from the
101 cervical environment (Ma et al., 2020). Moreover, we recently demonstrated that AcrF
102 in *E. coli* O157:H7 is non-functional due to a conserved insertion containing two stop

codons (Pugh et al., 2023), showing that within a species not all efflux pumps present are always functional.

The expression of RND efflux pumps is controlled by a complex network of positive and negative regulators including repressors of the TetR family. We previously reported the presence of the TetR family regulator EefR in four of ten *E. coli* strains studied (Colclough et al., 2019). The *eefR* gene was encoded alongside genes annotated as *eefA* and *eefB* that are predicted to encode a PAP and RND pump, respectively. This RND pump was first reported in *Enterobacter aerogenes* (now reclassified as *Klebsiella aerogenes*), where *eefA, eefB* and *eefC* were found in a three gene operon without the regulator, *eefR* (Masi et al., 2005). In that study, the *eefABC* operon was not expressed under laboratory conditions due to transcriptional silencing by H-NS (Masi et al., 2005), however overexpression was found to confer resistance to erythromycin (Masi et al., 2006). The EefABC RND pump has also been reported in *Klebsiella pneumoniae,* where it has been linked to virulence, as deletion of *eefA* was found to reduce both colonisation of the gastrointestinal tract and tolerance to low pH (Coudeyras et al., 2008).

In *E. coli* the *eefABC* operon is absent from the widely studied strain K-12 but has been reported in a single highly drug resistant environmental isolate*, E. coli* SMS-3-5. In this isolate the efflux pump was part of a larger gene cluster which included the regulator *eefR* and a major facilitator superfamily (MFS) efflux pump *eefD* (Fricke et al., 2008). To date, the wider prevalence of the *eefRABCD* operon across the diversity of the *E. coli* population and its function in *E. coli* are both still unknown. Here we demonstrate that: 1) EefABC is exclusively found in clinically relevant phylogroups of *E. coli* and is highly conserved, 2) homology modelling reveals that the pump has several distinctive structural features compared to other RND efflux pumps, and 3) neither EefABC, nor EefD transport clinically relevant antimicrobials but can transport dyes.

Materials and methods

Prevalence and conservation of *eefRABCD* across *E. coli* and related species

The *eefB* gene from *E. coli* SMS-3-5 (CP000970.1) was aligned to *E. coli* genomes (taxid 562) within the NCBI RefSeq Genome Database.

A total of 20,013 *E. coli* genome assemblies from 38 STs (Achtman scheme) (Wirth et al., 2006) were downloaded from Enterobase (Zhou et al., 2020) for the

determination of *eef* conservation across *E. coli*. Duplicates were removed based upon MASH (v 2.2.2) distances (Ondov et al., 2016). A custom blast database (BLAST v 2.10.0) (Camacho et al., 2009) was generated from the sequences of *eefR, eefA, eefB, eefC* and *eefD* of *E. coli* ATCC 25922 (CP009072.1) and used with ABRicate (v 0.9.8) (Seemann, 2017) to identify the presence and conservation of *eef* genes across the *E. coli* assemblies. Where an *eef* gene was found to be split across multiple contigs within an assembly, the assembly was removed from analysis as it was not possible to confirm from the assemblies alone whether this was due to a sequencing or assembly error, or the interruption of the gene.

The *eefRABCD* gene sequences from *E. coli* ATCC 25922 were also used to determine whether *Shigella* species encode the operon. Assemblies of *Shigella boydii* (*n* = 495), *Shigella dysenteriae* (*n* = 497), *Shigella flexneri* (*n* = 499), and *Shigella sonnei* (*n* = 500) were downloaded from Enterobase. As with *E. coli,* duplicates were identified using MASH and removed prior to running ABRicate. Assemblies containing *eef* genes split over multiple contigs were removed from the analysis as with *E. coli.* The *eefRABCD* genes from *E. coli* ATCC 25922 were also aligned to *Salmonella* (taxid 590), *Photorhabdus* (taxid 29487), *Yersinia* (taxid 629), *Serratia* (taxid 613), *Pseudomonas* (taxid 286), *Enterobacter* (taxid 547) and *Acinetobacter* (taxid 469) genomes to identify any homologs in related Gammaproteobacteria, however this was achieved using the NCBI RefSeq Genome Database.

The phylogenetic context of *eefRABCD*

Five assemblies of each *E. coli* ST (excluding ST84 where *n* = 2) and *E. fergusonii* were chosen at random to generate a phylogenetic tree. Assemblies were annotated using Prokka (v 1.14.6) (Seemann, 2014) with subsequent GFF files used as input for Roary (v 3.13.0) (Page et al., 2015). The core gene alignment produced by Roary was used to construct a GTR-gamma tree with 100 bootstraps using RaXmL (v 8.2.12) (Stamatakis, 2014). Trees were visualised and annotated using iTOL (Letunic and Bork, 2019).

Genomic context of *eefRABCD*

Due to limited availability of RefSeq genomes for some *E. coli* STs used in this work, seven were chosen at random. Reference sequences were downloaded from NCBI NC_004431.1 (ST73), NC_007946.1 (ST95), NZ_HG941718.1 (ST131),

NC_002695.2 (ST11), NC_000913.3 (ST10), NC_011751.1 (ST69), and NZ_CP035350.1 (ST617). The location of the *eef* operon was identified in the ST73, ST95, ST131 and ST11 and downloaded along with the flanking 10,000 bp. The homologous regions in ST10, ST69 and ST617 were also identified and downloaded. Alignments of the genomic regions in all seven reference sequences was performed using EasyFig (v 2.2.2) (Sullivan et al., 2011).

Genomes of additional bacterial species; *E. albertii* (CP070290.2), *E. marmotae* (CP056165.1), *K. aerogenes* (NZ_CP041925.1), *K. pneumoniae* (NC_016845.1), *Enterobacter vonholyi* (VTUC01000001.1), *Enterobacter dykesii* (VTTY01000003.1), *Enterobacter wuhouensis* (SJOO01000006.1), *Enterobacter kobei* (KI973153.1), *Enterobacter chengduensis* (CP043318.1), *S. boydii* (CP026836.1), *S. dysenteriae* (CP026774.1), *S. flexneri* (AE005674.2) and *S. sonnei* (CP055292.1) were downloaded from NCBI. The visualisation of the genomic context of the *eef* operon and homologs was achieved using EasyFig. Mapping of insertion sequences in *S. dysenteriae* was done using ISEScan (v 1.7.2.3) (Xie and Tang, 2017).

Homology modelling and structural analysis

Multiple sequence alignments (MSA) were prepared using MAFFT and NJ/UPGMA phylogeny algorithms as implemented in MAFFT v.7 server (https://mafft.cbrc.jp/) (Katoh et al., 2019). Phylo.io was used for phylogenetic guide tree visualisation (Robinson et al., 2016). Structural annotations of the MSA sequences were done with Espript 3 (http://espript.ibcp.fr) (Robert and Gouet, 2014).

For homology modelling, I-TASSER (Yang et al., 2015) was used in manual mode with assignment of templates and structural alignment, supplemented by SWISS-MODEL (Waterhouse et al., 2018). The following structural templates have been used for the specific protein modelling. EefA modelling: MexA (Uniprot P52477) 2V4D.pdb (Symmons et al., 2009); AcrA (Uniprot P0AE06) 5V5S.pdb (Wang et al., 2017). AcrA was used as a template due to the smaller gaps in the alignment and better quality of available full-length template. EefB modelling: MexB (Uniprot P52002) 3W9I.pdb (Nakashima et al., 2013); AcrB (Uniprot P31224) 2GIF.pdb (Seeger et al., 2006). EefC modelling: TolC (Uniprot P02930) 1EK9.pdb (Koronakis et al., 2000); OprM (Uniprot Q51487) 4Y1K.pdb (Monlezun et al., 2015); OprJ (Uniprot Q51397) 5AZS.pdb (Yonehara et al., 2016); OprN (Uniprot Q9I0Y7) 5AZO.pdb, 5AZP.pdb

(Yonehara et al., 2016); 5IUY.pdb (Ntsogo Enguéné et al., 2017). EefD modelling: EmrD (Uniprot P31442) 2GFP.pdb (Yin et al., 2006); MdfA (Uniprot P0AEY8) 4ZP0.pdb (Heng et al., 2015).

For the models of the protein oligomers and the complete EefABC tripartite pump, rigid-body structural docking of the homology models was used guided by the available cryo-EM structure (5O66.pdb; (Wang et al., 2017)), the results of which were cross-validated manually and obvious steric clashes removed using Coot (Emsley et al., 2010). Additional structural analysis and visualisation was performed with Pymol (PyMOL Molecular Graphics System, Version 1.71 Schrödinger, LLC).

Strains, plasmids and culture conditions

All strains used in this work are listed in Table S1. Strains were grown in lysogeny broth (LB, Merck) at 37˚C with aeration unless stated otherwise.

Cloning of *eefABCD*

The *eefABC* operon was amplified from the chromosome of *E. coli* ATCC 25922 (NCTC 12241) using Q5 polymerase (New England Biolabs) and primers which incorporated the *Nde*I and *Xho*I restriction sites (Table S2). The amplicon was cloned into both pET21a (ampicillin resistant) and pET24a (kanamycin resistant) plasmids (Invitrogen) which are identical aside from their resistance cassette. No IPTG induction was used in this work.

The *eefD* gene was amplified from the *E. coli* ATCC 25922 chromosome using Q5 (New England Biolabs) and primers that incorporated the *ApaLI* and *PstI* restriction sites (Table S2). The amplicon was then cloned into pACYC177 (ATCC).

Deletion of *eefABC* and *eefD* in *E. coli* ATCC 25922

Deletion of the genes encoding the RND system of the *eefRABCD* operon was achieved by homologous recombination (Kim et al., 2014). However, due to the size of the *eefABC* operon, first *eefB* was interrupted, followed by the remaining *eefA* and *eefC* genes. An *eefD* knockout was generated independently using the same method. All primers are listed in Table S2.

Determination of minimum inhibition concentration (MIC) of antimicrobials, metals, and dyes

Bacterial susceptibility to a range of antimicrobials and dyes was determined using the agar doubling dilution method described by the Clinical and Laboratory Standards Institute (CLSI, 2020). *E. coli* ATCC 25922 was used as a control to confirm antimicrobial efficacy in line with EUCAST guidelines (EUCAST, 2021). For the susceptibility to metals, bile salts and polyamines, a broth microdilution method was used (EUCAST, 2021).

Accumulation and efflux of ethidium bromide

Ethidium bromide (EtBr) accumulation was measured as previously described (Smith and Blair, 2014). Briefly, EtBr was added to cells and the increase in fluorescence was measured over time.

Efflux activity was also assessed as previously described (Smith and Blair, 2014). Here, cells were incubated in the presence of EtBr and carbonyl cyanine m-chlorophenyl hydrazone (CCCP) until fluorescence saturation was reached. Re-energisation was achieved with glucose and the rate of reduction in fluorescence was measured.

Results

The *eefRABCD* operon is present in *E. coli* phylogroups associated with infection

To first understand how widespread the *eefRABCD* operon is across *E. coli,* the NCBI RefSeq Genome Database was utilised. The *eefB* gene from *E. coli* SMS-3-5 was aligned to genomes belonging to *E. coli* (taxid 562) using NCBI nucleotide blast. The top 100 matches in the *eefB* alignment had 100% sequence coverage and ≥97.7% sequence identity (data not shown), suggesting that the *eefB* gene was present more widely across *E. coli*. One of the strains found to possess *eefB* was the well characterised strain ATCC 25922. The *eef* operon in both *E. coli* SMS-3-5 and ATCC 25922 were found to be very similar and as a result, the operon from ATCC 25922 was used as the reference in future work.

Next we looked at the distribution of *eefRABCD* across *E. coli*. 20,013 assemblies were downloaded from the Enterobase database representing 38 STs. Using MASH distances, 766 assemblies were identified as duplicates and removed from the data set. The presence of *eefR, eefA, eefB, eefC* and *eefD* in the remaining

19,247 assemblies was determined using ABRicate (Table S3). Interestingly there was a clear divide between phylogroups where the *eefRABCD* operon was identified and those where it was completely absent (Fig. 1). The operon was not found in ST groups belonging to phylogroups A, B1 and C, which are more traditionally classified as environmental isolates. However, the *eefRABCD* operon was present in all ST groups of phylogroups B2, F and G, which are strongly associated with human infection and MDR. Notably, a single ST within phylogroups D and E lacked the operon, ST69 and ST182 respectively. Despite these two exceptions, a distinct divide between the A-B1-C and B2-D-E-F-G clades was identified, implying a clear evolutionary relationship.

The *eefRABCD* operon is highly conserved across phylogroups B2, D, E, F and G

In *E. coli* assemblies positive for an *eefRABCD* gene, the entire operon was always present, with gene sequences highly conserved. Using *E. coli* ATCC 25922 as a reference, gene coverage for each component of the *eef* operon was always greater that 98% (Table S6). The nucleotide percentage identity varied between genes and across phylogroups though despite subtle variations, *eefR, eefA, eefB* and *eefD* averaged >99% overall. Interestingly, *eefC,* the gene coding for the OMF, was marginally more variable than the four other genes within the operon (Table 1).

To contextualise the extent of *eefRABCD* conservation, the conservation of *acrA, acrB* and *tolC*, which encode the critically important RND efflux pump AcrAB-TolC in *E. coli*, was determined. In general, the nucleotide identity of the *eef* operon was conserved to a similar level as *acrAB-tolC*, all genes (excluding *eefC* in ST354) were >97% identical to those in *E. coli* ATCC 25922 (Table 1). However, conservation of both the *acrAB-tolC* and *eefABC* RND systems varied slightly at both the ST and phylogroup level. This suggests a strong selection pressure on the operon, indicating an important biological function.

Phylogroup B2 contains many clinically significant *E. coli* STs including the MDR ST131. Across phylogroup B2, *eefB was* generally more conserved than *acrB* with the *eef* genes having higher homology to the reference than the *acrAB-tolC* genes. However, in phylogroups D, E and F the opposite was seen as the *acrAB-tolC* genes were more conserved than the *eefRABCD* genes. Though only one phylogroup

G ST was used in the analysis, in ST738 *eefR, eefC* and *eefD* had the highest homology to the reference, with *acrA, acrB, eefA* and *eefB* all equally conserved.

Phylogroups A, B1 and C have highly conserved *sapF-fabI* intergenic region in place of *eefRABCD*

While *eefRABCD* was found to be highly conserved across phylogroups B2, D, E, F and G, it was completely absent from phylogroups A, B1 and C. To explore this further RefSeq genomes of strains with and without the operon were downloaded from NCBI and aligned using EasyFig (Fig. 2). In assemblies encoding the *eef* operon, it was always found at the same genomic location; that is between the essential gene *fabI* and the non-essential gene *sapF*. FabI is an enoyl- [acyl-carrier-protein] reductase that is involved in fatty acid production (Bergler et al., 1994), while SapF is a putrescine export protein belonging to the *sapBCDF* system (Sugiyama et al., 2016). Interestingly, in assemblies where the *eefRABCD* operon was absent, a hypothetical gene, *ycjD,* was annotated as present in the same genomic location. The *ycjD* open reading frame (ORF) was 354 nucleotides long and ran in the opposite orientation to *eefRABCD*.

A larger alignment of the 97 assemblies from phylogroups A, B1 and C (and ST69 and ST182) used to construct the phylogenetic tree demonstrated that the *ycjD* hypothetical gene is highly conserved across phylogroups A, B1 and C (Supplementary Data file). The alignment of the *sapF-ycjD-fabI* region also demonstrated that ST69 and ST182 possess an intergenic region highly homologous to the STs belonging to phylogroups A, B1 and C, though subtle differences were present (Fig. 3). The open reading frame annotated as *ycjD* in K-12, is 12 nucleotides longer in both ST69 and ST182 assemblies (Fig. 3). Taken together, these findings suggest that *eef* has been lost from *E. coli* on at least two independent instances.

Distribution of *eefRABCD* in Gram-negative bacteria

To see if the *eefRABCD* operon was present more widely across *Escherichia* species, the ATCC 25922 *eef* operon was aligned to sequences from the *Escherichia* genus (taxid 561), with *E. coli* (taxid 562) sequences excluded, using the NCBI RefSeq database. Only *Escherichia marmotae* and *Escherichia albertii* were found to possess the *eefRABCD* operon (Fig. S1) while *E. fergusonii, Escherichia ruysiae*, *Escherichia vulneris, and E. hermannii* did not encode it.

Due to the genetic similarity of *Shigella* and *E. coli* (van den Beld and Reubsaet, 2012, Brenner et al., 1972), the presence of the *eef* operon across *Shigella spp*. was investigated. Gene fragments were detected in a small number of *S. boydii* and *S. flexneri* assemblies but only *S. dysenteriae* was found to consistently possess the operon (Fig. S2, Table S5). However, whilst genes belonging to the *eefRABCD* operon were detected in 363 of the 486 *S. dysenteriae* assemblies included in this study, only a single assembly was positive for *eefR.* Moreover, whilst *eefA* was present in all assemblies possessing the operon, sequence coverage averaged at 24.6%. Assembly annotation identified the presence of an insertion sequence in the place of *eefR* and *eefA* (Fig. S3), explaining the absence and truncation of *eefR* and *eefA* respectively, across the *S. dysenteriae* assemblies. In comparison, no insertion sequences were detected within the *eef* operon of *E. coli* ATCC 25922, nor in the 10,000 base pairs up- or down-stream of the operon.

As the EefABC efflux pump has been reported in both *K. aerogenes* and *K. pneumoniae* (Coudeyras et al., 2008, Masi et al., 2005) with differing operon architecture (Fig. S4), the presence of the operon across related Gammaproteobacterial genera was determined. Only low homology orthologs of EefA and EefB were identified in *Yersinia* and *Serratia* while no conserved homologues were identified in *Salmonella*, *Acinetobacter*, *Pseudomonas* or *Photorhabdus* (Table S6). In *Enterobacter,* conservation of the operon differed between species (Fig. S5). Some species, such as *E. chengduensis*, encoded *eefRABC,* but not *eefD*, however sequence identity was only 73% (blastn) indicating that the *E. coli* and *Enterobacter eef* operons are not homologous, and instead the association is likely due to historical literature and gene nomenclature.

Sequence and structural prediction analysis reveals unique features of the EefABC pump

The *eefABC* operon appears to encode a tripartite RND-efflux pump similar to the AcrAB-TolC assembly in *E. coli*. As there is no experimental structural information currently available on any of the components of EefABC, we conducted sequence analysis to identify potential structural templates and subsequently performed homology modelling using the highest scoring templates.

Comparison with other OMFs of known structure, revealed the closest relatives within the wider OMF-family to be the *Pseudomonas* proteins OprM and OprJ (Fig.

S6), followed by CusC, and hence OprM/J were used as structural templates for homology modelling (Fig. S7 and S8 and supplementary text). Structural alignment of EefC with OprM/J produced an alignment with very few gaps allowing for direct mapping of the aligned sequences of EefC onto OprM/J. EefC has significantly shorter extracellular loops in comparison to TolC, in particular the L2 loop which occludes TolC opening (Vaccaro et al., 2008), suggesting a more-open state of the EefC channel (Fig. 4a). Such OMF loops are prominent targets of protective antibodies (Domínguez-Medina et al., 2020) and the non-protruding loops may serve to avoid antibody restriction and LPS occlusion. Additionally, the N-terminal tail of EefC is 30 residues longer than that seen in TolC, making it more similar to the structure of OprN/J (Fig. S8 and Fig. S9).

The comparison of the gating-loop regions in EefC and TolC reveals that they are highly divergent, indicating a locking mechanism markedly different from that observed in TolC, and hence EefC is unlikely to function with any PAPs that normally pair with TolC, and likely only interacts with its cognate PAP EefA. Specifically, several key residues in the helix7/helix8 hairpin of TolC, responsible for gating the TolC channel are different in EefC, e.g. the R367 (TolC), forming part of the so-called “primary gate” that seals the TolC channel by binding to the conserved D153 (TolC; D206 in EefC) and thus anchoring to the helix 4 (Andersen et al., 2002a, Bavro et al., 2008), in EefC is substituted by a small non-polar residue G412, making such interaction impossible (Fig. S8).

In addition, the electrostatic properties of the EefC channel are predicted to be dramatically different to TolC and other OMFs, which is likely to impact heavily the ion and drug selectivity of the channel (Marshall and Bavro, 2020). Firstly, the so-called “secondary gate” of the channel, formed of the prominently conserved double aspartate ring (D371; D373) which forms the basis of cation-selectivity in TolC (Andersen et al., 2002b, Schulz and Kleinekathöfer, 2009), is fully absent in EefC and instead is substituted by small-hydroxylated residues (T416 and T419), a feature which appears to be unique to EefC. Secondly, EefC possesses additional bulky aliphatic residues L415 and L418, which effectively hydrophobically seal the periplasmic end of the channel (Fig. S8).

The PAP EefA was analysed in a similar way, revealing a sequence identity of 53.85% to the *Pseudomonas* MexA; but only 50.13% against the *E. coli* AcrA. Due to the availability of more complete full-length experimental templates, EefA homology

models were created using both MexA (2v4d.pdb; (Symmons et al., 2009)) and AcrA (5v5s.pdb; (Wang et al., 2017)) as templates, (utilising both I-TASSER and Swiss-model tools) (Fig. 4b; Fig. S10). While neither of the models delivered the same level of confidence as those for EefC, the alignments with the known PAP structures are readily interpretable, allowing identification of the protein features. A detailed discussion of the EefA structure is given in the Supplementary text with the major findings summarised here.

Alignment of EefA and AcrA results in a direct amino-acid match with only 2 gaps in the alignment, one in the unstructured N-terminal tail, and another at position Q221 (EefA), which has a 4-residue-long deletion relative to AcrA (Fig. 4b; Fig. S10**)**. This region corresponds to the C-terminal end of α-helix 3, which is flanking the β-barrel domain in PAPs, however it is not predicted to affect RND-binding (McNeil et al., 2019). Despite overall similarity with MexA/AcrA there are distinctive differences in the organisation of the EefA, notably in its α-hairpin domain (Fig. 4b; Fig. S11). While the RLS(D) motif, which is thought to be critical for PAP-OMF interaction (Kim et al., 2010, Song et al., 2014, Alav et al., 2021) appears to be preserved in EefA, (R120/L124/S131/D136), there are a number of significant changes in the adjacent residues, notably R123 (K131 in AcrA); V125 (L133); D134 (E142), which would likely result in steric clashes that would preclude direct compatibility with TolC, and such residues are likely to be playing a discriminatory role engaging with EefC (Fig. S11).

The RND-transporter component, EefB, is a large integral membrane protein, predicted to form a functional trimer similar to other transporters in the family (Fig. 4c). Alignment of the EefB with other RND transporters of known function including AcrB (*E. coli*); MtrD (*N. gonorrhoeae*); MexB (*P. aeruginosa*); CusA (*Campylobacter jejuni*) and AdeB (*A. baumanii*) revealed that they are highly similar, with EefB being the most similar to AcrB and MexB (57.0% and 56.8% identity respectively), while CusA and MtrD being most divergent (Fig. S12). The MexB structure was used as a template to generate a high-fidelity homology model of EefB, as there were fewer gaps in the alignment (0.5% vs 0.7% for AcrB) and slightly higher residue overlap with EefB (1038 *vs* 1033 residue overlap respectively).

Overall, the alignment of EefB with MexB (and AcrB) produces very few gaps (the longest is 2 residues long), allowing for unequivocal attribution of secondary structure elements (Fig. S12). As can be seen from the side-by-side comparison of the EefB and AcrB, both present a virtually identical architecture (Fig. S13), with the

only notable differences being the shortened loop connecting the TM helices α16 and α17 in EefB (residues 498-507); and the shorter C-terminal tail. Consistent with this, the critical proton-relay residues found in MexB D407, D408, K939 and T976 (Guan and Nakae, 2001) are conserved in EefB (D408, D409, K935 and T972 respectively) and the predicted structures of the access and distal binding pockets of EefB suggest a closer relation to the MexB/AcrB than to MexY type transporters.

The last member of the operon is EefD, a member of the MFS family of efflux pumps that according to the Transporter Classification Database (TCDB) belongs to the 2.A.1.2 group of MFS transporters, related to MDR-function, which also includes Bcr/CflA, EmrD and MdfA (Reddy et al., 2012, Quistgaard et al., 2016) (Fig. S14). This resulted in high-confidence homology models (I-TASSER C-score = 1.82), revealing a classic 6+6 transmembrane helical arrangement closely related to the general topology of MdfA, although with significant differences in the substrate binging cavity, suggesting the substrate range will be notably different between EefD and MdfA (Fig. 4d; Fig. S15) (Adler and Bibi, 2002, Heng et al., 2015, Nagarathinam et al., 2018). These discrepancies make predictions of the possible substrates of EefD problematic, although some overlap with MdfA can be expected (Lewinson et al., 2003), e.g. lipophilic cations such as ethidium.

We used the homology models described above to dock the components into a complete tripartite pump using the cryo-EM structures of AcrAB-TolC (Wang et al., 2017) as a guide. EefABC can indeed be assembled using the same architecture with minimal steric clashes, as can be seen in Fig. 5.

E. coli **EefABC is not a drug transporter**

The high levels of conservation of the *eefRABCD* operon within clinically relevant lineages of *E. coli* suggests it has an important biological function and its substrate profile may differ from other *E. coli* RND pumps due to its unique structure. Antimicrobials are known substrates of RND efflux pumps such as AcrB, and overexpression of RND pumps can confer MDR in both the laboratory and clinic. Therefore, the effect of *eefABC* and *eefD* expression on *E. coli* susceptibility to a range of antimicrobials and dyes was determined. Due to the previously mentioned homology between *E. coli* ATCC 25922 and *E. coli* SMS-3-5, and the well characterised antimicrobial susceptibility profile of ATCC 25922 this strain was used for EefABC characterisation experiments.

Deletion of *eefB* in *E. coli* ATCC 25922 did not increase susceptibility to any antimicrobial or dye tested. Subsequent inactivation of *eefA* and *eefC* to give an *eefABC* knockout also had no effect on the susceptibility of *E. coli* ATCC 25922 to antibiotics (Table 2).

As loss of EefABC function did not alter the drug susceptibility profile of ATCC 25922, *eefABC* was cloned into the pET21a and pET24a plasmids and expressed in *E. coli* MG1655, which does not naturally encode the system. Expression of *eefABC* in *E. coli* MG1655 did not alter susceptibility to any of the antimicrobials tested. AcrB is the dominant RND pump in *E. coli* and this can mask phenotypic changes associated with other efflux systems so it was deleted. However, expression of *eefABC* in the absence of *acrB* still did not reveal any significant changes in susceptibility but did result in a decrease in susceptibility to ethidium bromide and rhodamine 6G suggesting these dyes can be transported by the pump (Table S7).

The MFS pump EefD was also characterised, however both inactivation in *E. coli* ATCC 25922 (loss of function) and expression in *E. coli* MG1655 (gain of function) had no effect on antimicrobial and dye susceptibilities. Expression of *eefD* alone in MG1655 Δ*acrB* was also not found to affect MICs.

A subset of the RND pumps (including CusABC) are known to pump metal ions and form the subfamily of heavy metal efflux (HME)-pumps (Gupta et al., 1999, Long et al., 2012, Lecointre et al., 1998, Klenotic et al., 2020). Due to the structural similarity of EefC to CusC, the susceptibility to heavy metals was also measured. Yet when either *eefB* or *eefABC* were deleted from ATCC 25922, the susceptibility to cobalt, nickel, zinc and iron were not significantly different to the wild-type ATCC 25922.

EefABC can export ethidium bromide

To confirm the cloned EefABC pump is functional, and that ethidium bromide is a substrate, the intracellular accumulation and efflux rate of ethidium bromide was measured. Expression of EefABC with EefD in the absence of AcrB significantly decreased intracellular accumulation of EtBr and increased the rate at which EtBr was pumped out of cells (Fig. S16 and S7). In addition, deletion of *eefD* significantly slowed efflux of ethidium bromide, which together, shows this inner membrane pump is needed for transport of the substrate across the inner membrane.

Discussion

The number of RND pumps present in different Gram-negative bacterial species varies and there is increasing evidence that prevalence of RND efflux pumps can also vary between isolates of the same species (Darby et al., 2023, Leus et al., 2020, Nowak et al., 2015, Nemec et al., 2007, Wieczorek et al., 2013). Yet it is still broadly assumed that all *E. coli* isolates possess six RND efflux pumps despite recent work showing that not all six pumps are always functional (Anes et al., 2015, Pugh et al., 2023). Here we further demonstrate that the assumption of six RND pumps is inaccurate; STs belonging to the phylogroups of *E. coli* that are most commonly associated with invasive infection (B2, D, E, F and G) encode a seventh, highly conserved RND pump operon (*eefRABCD*), while the operon was completely absent from phylogroups A, B1 and C, which are generally associated with environmental or commensal lifestyles. The level of conservation suggests that EefRABCD has a biologically important function resulting in a high degree of selection pressure while the distribution in phylogroups commonly associated with infection suggests the system could have a role in infection or survival in the host environment.

Across phylogroups A, B1 and C a 354-nucleotide ORF is found in place of *eefRABCD.* This putative gene *ycjD* runs in the opposite orientation to *eefRABCD* and is highly conserved between STs of these phylogroups. This putative gene is also present in ST69 and ST182 which belong to phylogroup D and E respectively though in these two STs a 12-nucleotide insertion is present at the 3' end of the gene. Studies from another group support the hypothesis that transcriptional activity is happening at this ORF as public data from their transcriptional start site, term-seq and ribosomal profiling studies show antisense transcriptional and translational activity at the 5' end of *ycjD* gene (Adams et al., 2021, Thomason et al., 2015, Weaver et al., 2019). It is worth noting, that while *ycjD* is indeed coding a domain of unknown function (DUF), the UniproKB AlphaFold structural prediction for it strongly suggests that it is linked to DNA modification/restriction, as it belongs to the endonuclease/DNA methylase fold.

The EefABC efflux pump was first described in the opportunistic human pathogen *K. aerogenes* and subsequently in *K. pneumoniae*, though operon structure differs between the two *Klebsiella* species (Coudeyras et al., 2008, Masi et al., 2005). In *K. aerogenes* overexpression of *eefABC* decreased susceptibility to erythromycin and ticarcillin (Masi et al., 2006) whilst overexpression of *K. pneumoniae eefA* and *eefB* in an *E. coli* (*ΔacrAB, ΔydhE*) background decreased susceptibility to oxacillin,

erythromycin, novobiocin, acriflavine, ethidium bromide and cholate (Ni et al., 2020). However, in this study, neither gain nor loss of *eefABC* function from *E. coli* altered susceptibility to any antibiotics tested, though over-expression did decrease susceptibility to ethidium bromide. Moreover, expression of *eefABC* (in the absence of AcrB) decreased intracellular accumulation of ethidium bromide demonstrating that the pump is functional. No effect was seen upon inactivation of the pump but this is likely because AcrB was still present which would mask any effect. However, deletion of the inner membrane component *eefD* alone, caused significantly slower EtBr efflux even in the presence of AcrB suggesting it has a significant role in transport of substrates across the inner membrane. The apparent difference in substrate profile between *Klebsiella* and *E. coli* could be due to differences in nucleotide sequence or operon structure subsequently altering the regulation of the operon or function of the orthologous protein. In *K. aerogenes* the operon has been demonstrated to be H-NS silenced in laboratory conditions (Masi et al., 2006, Masi et al., 2005), though due to the genomic location of the *eef* operon in *E. coli* and the resulting proximity of the *eefR* promoter to that of *fabI* which encodes an essential enoyl- [acyl-carrier-protein] reductase it is unlikely that the operon is silenced *in vivo* in *E. coli.*

When looking across the *Escherichia* genus, *eefRABCD* was only present in *E. marmotae* and *E. albertii,* though the operon was identified in *S. dysenteriae,* which has high genetic similarity to *E. coli.* Here, *eefR* was generally absent due to the presence of an insertion sequence in place of *eefR* and *eefA.* A further insertion sequence was identified downstream of *eefD.* It has been shown previously that *Shigella* species have higher numbers of insertion sequences when compared to other Gammaproteobacteria such as *E. coli* (Touchon and Rocha, 2007). High numbers of insertion sequences is linked to recent host-specification as the integration of IS elements into the genomes is associated with early stage genome degradation (Moran and Plague, 2004, Hawkey et al., 2020, Yang et al., 2005). The presence of an insertion sequence in place of *eefR* and *eefA* and downstream of *eefD* may therefore suggest the operon is in the process of being lost or degraded in *S. dysenteriae*. Many Gram-negative bacteria including *Salmonella* sp. (taxid 590) did not encode an ortholog of the Eef system but orthologs of EefA and EefB were identified in *Serratia* and *Yersinia* species, suggesting related systems could be dispersed through the Gammaproteobacteria.

Overall, our analysis of EefABC revealed unexpected similarities to the tripartite pumps OprM-MexAB and OprJ-MexCD in *Pseudomonas* (Masuda et al., 2000), rather than to any endogenous *E. coli* paralogues, which may suggest that these pumps have been acquired *via* a lateral gene transfer event, similar to some other plasmid encoded efflux/secretion systems e.g. the EAEC virulence plasmid pAA2 (Imuta et al., 2008).

Given the close similarities of the EefABC system to both MexAB-OprM and MexCD-OprJ and based on previous profiling of these *Pseudomonas* pumps (Masuda et al., 2000), we initially hypothesised that these pumps could have overlapping substrate ranges. However, despite the overall similarity, our tests did not reveal any direct effect of EefABC on transport of antimicrobials, suggesting alternative function.

As *E. coli* also possesses HME-RND systems such as CusABC we also investigated the possibility that EefABC may be involved in metal ion tolerance, but we were unable to identify a metal ion substrate for the efflux pump (Table 8). As mentioned above, the EefC lacks the double aspartate "gates" which have been implicated in the coordination of metal ions in the case of TolC (Higgins et al., 2004, Bavro et al., 2008) (Fig. S8). This alone however could not be used to rule out ion-efflux function, as it has previously been noted (Kulathila et al., 2011) that CusC does not (Alcalde-Rico et al., 2016)by itself show any specific features associated with monovalent ion selectivity, with that function rather being associated with the RND transporter, in that case CusA. Indeed, while CusA presents a clear relay of methionine clusters to bind and export Cu(I) and Ag(I) ions (Delmar et al., 2013), there is no identifiable pattern of such residues in the case of EefB. On the other hand, divalent-ion specific pumps such as ZneA, have more restricted binding sites formed from negatively charged residues (Pak et al., 2013), but again, the residues participating in ion coordination are not conserved in EefB.

Aside from conferring resistance to antimicrobials, RND efflux pumps have myriad other physiological roles involving export of both exogenous and endogenous substrates (Henderson et al., 2021). For example, RND efflux pumps in various Gram-negative bacteria have been linked to virulence (Wang-Kan et al., 2017, Nishino et al., 2006). While we have so far been unable to assign clear function to EefRABCD, given the striking distribution and conservation of *eefRABCD* only in phylogroups of *E. coli* associated with disease it is possible that this pump has roles associated with virulence or survival in the host environment and work is continuing to assign biological function.

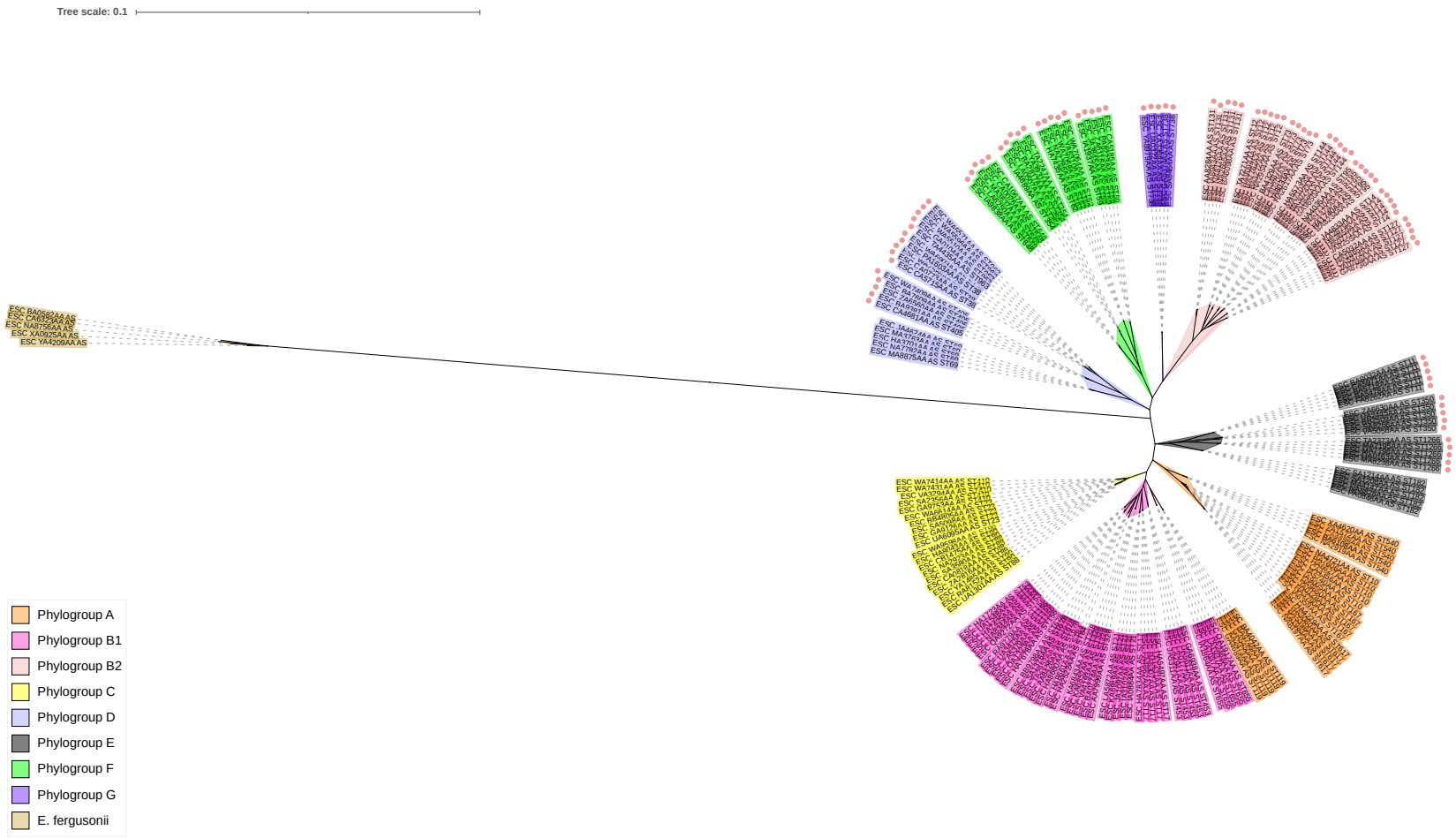

Figure 1. Phylogenetic structure of the assemblies used in this analysis and the distribution of *eefRABCD*

The tree was created using five assemblies per ST (exception ST84 *n=2*). Assemblies were chosen randomly. The tree was rooted using five *E. fergusonii* assemblies. The leaves are annotated with the ST group of the assembly and coloured coded by phylogroup, colours used to highlight phylogroups are shown in the legend on the bottom left corner. STs positive for the *eefRABCD* operon are marked with ●. Tree was generated using RaXmL and visualised on iTol.

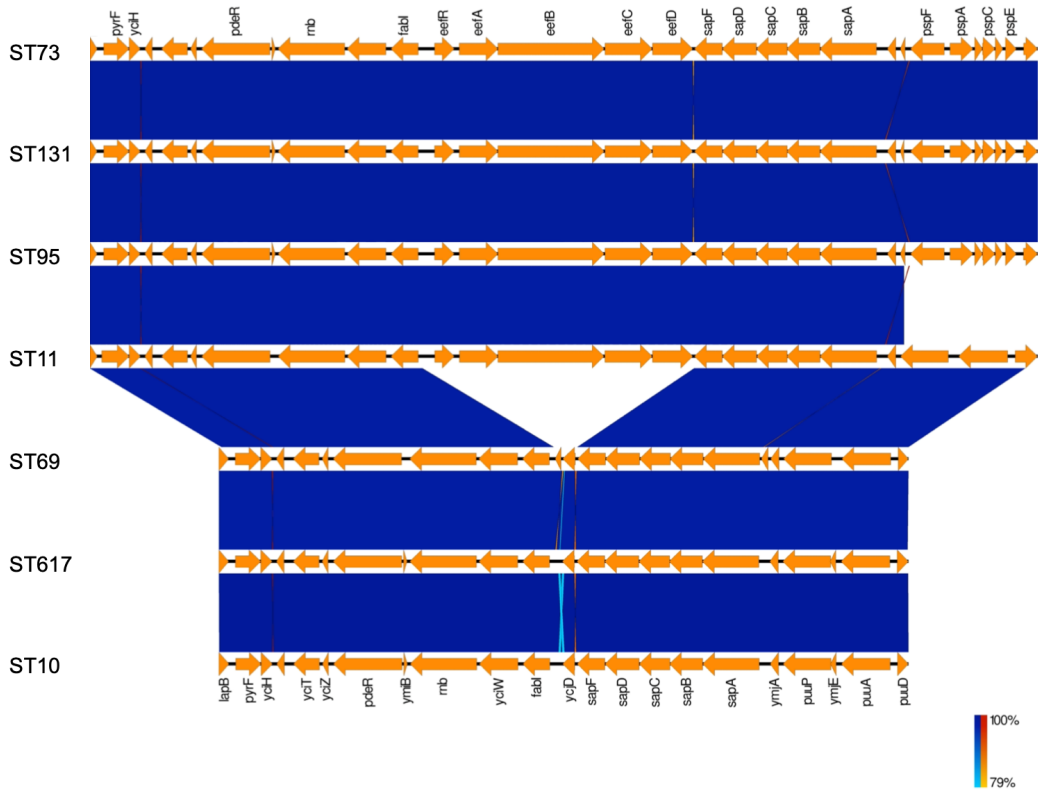

Figure 2. The genomic region of strains with and without *eefRABCD*. EasyFig alignment of the *eef* region in ST73 (B2), ST131 (B2), ST95 (B2), ST11 (E), ST69 (D), ST617 (A), ST10 (A). The *eefRABCD* operon was consistently identified between *fabI* and *sapF.*

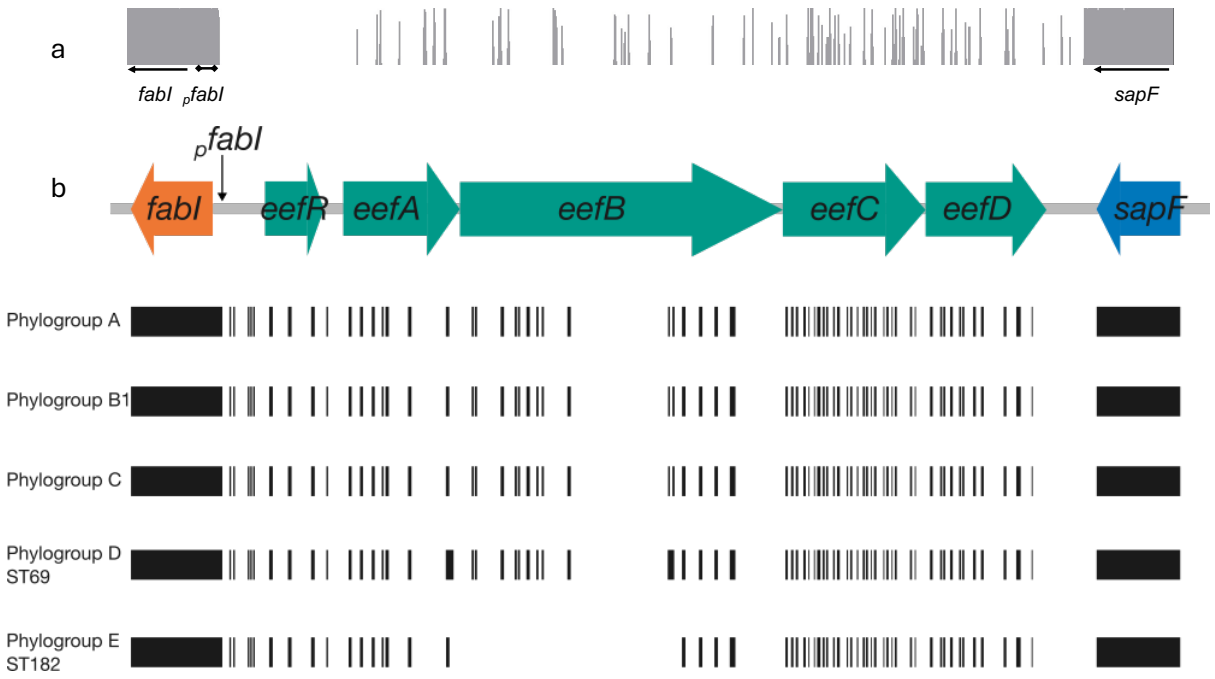

Figure 3. Diagrammatic representation of *eefRABCD* homology in strains that lack the operon.

In *E. coli* ST positive for *eefRABCD*, the operon was consistently located between *fabI* and *sapF*. To determine whether *E. coli* assemblies that lacked the operon had conserved regions of the *eef* operon, the equivalent *fabI-sapF* intergenic region of assemblies negative for *eefRABCD* were aligned with *fabI-eefRABCD-sapF* from *E. coli* ATCC 25922. **A) Consensus identity of the *fabI-sapF* intergenic region in *E. coli* that lack the *eef* operon.** Between all STs, and despite phylogroup, the *fabI-sapF* intergenic regions aligned to the *eefRABCD* operon in a highly conserved manner. **B) Cartoon representation of operon fragment conservation in *E. coli* ST that lack the *eef* operon.** Closer inspection of the alignment (Supplementary file 1) found that phylogroups A, B1 and C had almost identical *fabI-sapF* intergenic regions, while ST69 and ST189 had marginally different homology patterns. Taken together, these data suggest the operon may have been lost in up to three independent events.

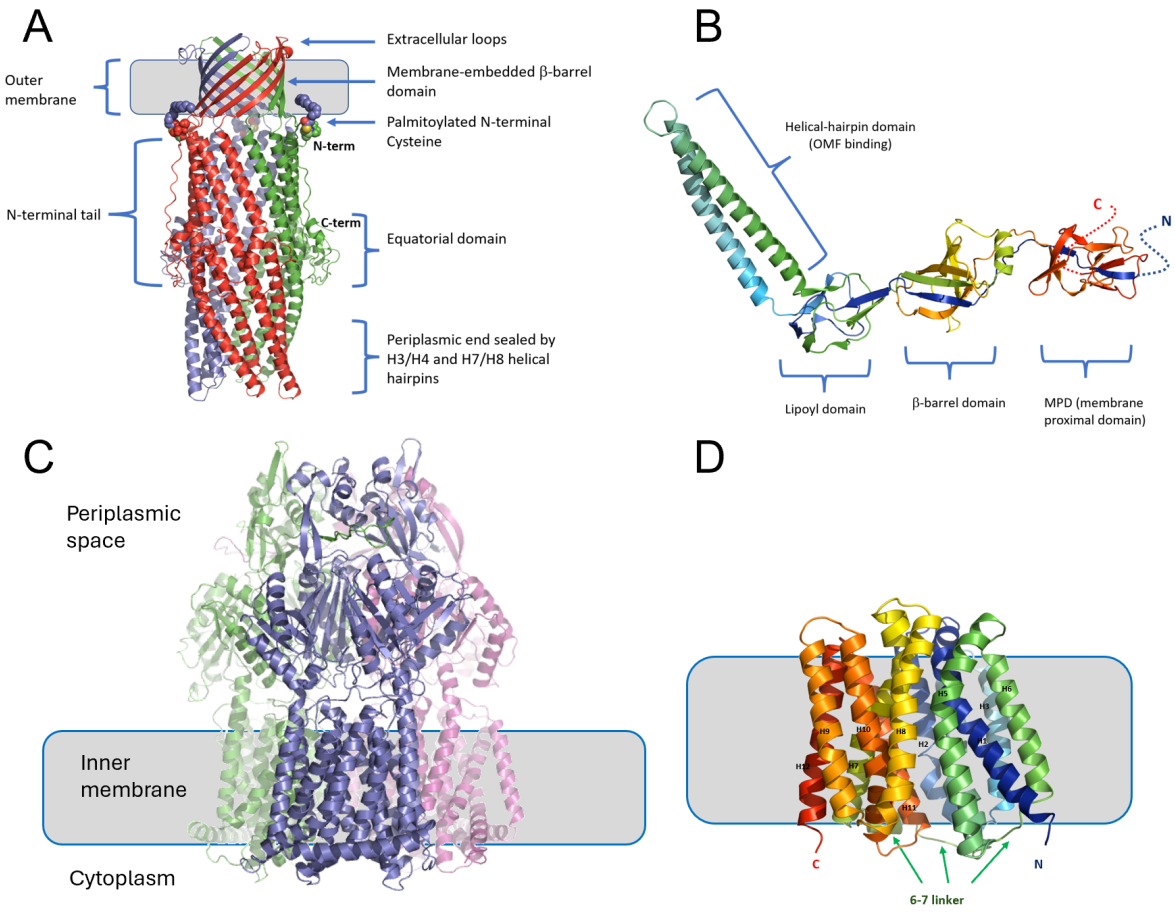

Figure 4. Multi-panel representation of the Eef components. Not to scale.

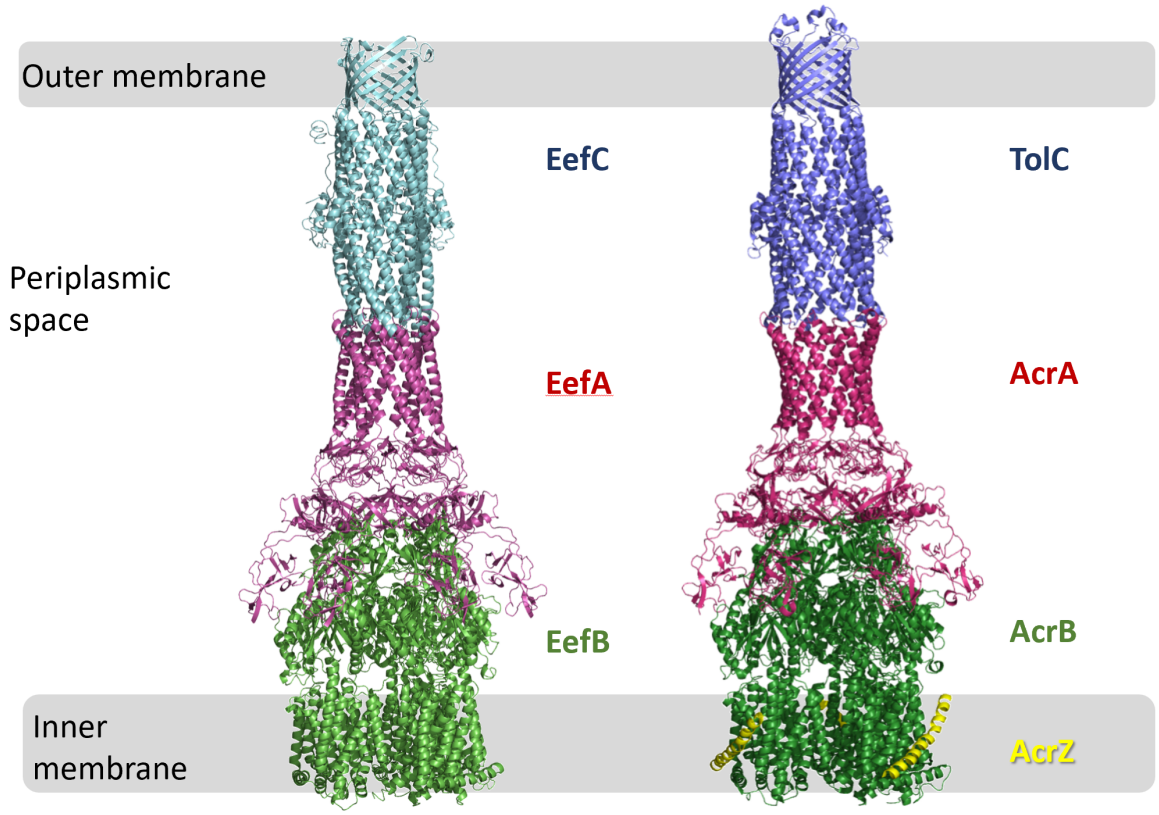

Figure 5. Comparison of the predicted structures of the assembled EefABC and the experimental cryo-EM structure of AcrABZ-TolC (based on 5066.pdb; (Wang et al., 2017))

Table 1. Conservation of *acrAB-tolC* and *eefRABCD* across *E. coli* phylogroups

Phylogroup	ST group	Mean nucleotide identity to *E. coli* ATCC 25922 (%)							
		eefR	*eefA*	*eefB*	*eefC*	*eefD*	*acrA*	*acrB*	*tolC*
B2	ST12	99.82	99.98	99.68	99.48	99.40	99.83	98.70	98.95
	ST127	99.12	99.64	99.61	99.42	99.48	99.75	99.46	98.57
	ST131	99.65	99.54	99.58	99.13	99.48	99.75	99.49	98.83
	ST14	98.91	98.49	99.41	99.12	99.57	99.83	99.70	98.83
	ST144	99.65	99.91	99.74	99.70	99.74	99.83	99.71	98.79
	ST73	100.00	100.00	100.00	100.00	99.97	100.00	100.00	99.93
	ST95	99.82	99.91	99.77	99.45	99.48	99.83	98.73	99.05
D	ST38	99.12	98.65	98.42	97.39	98.96	99.25	98.69	98.43
	ST405	99.45	98.32	99.00	99.20	99.14	99.33	98.67	98.05
	ST69	-	-	-	-	-	99.16	98.76	98.29
	ST963	99.12	98.66	98.42	97.38	98.97	99.25	98.70	98.65
E	ST11	99.12	97.95	98.06	97.53	98.37	99.33	98.53	98.52
	ST182	-	-	-	-	-	99.17	98.64	98.38
	ST350	99.47	98.03	98.38	97.62	98.25	99.08	98.70	98.44
	ST1266	99.47	98.22	98.46	97.57	98.19	99.08	98.76	98.65
F	ST354	98.93	98.84	99.52	97.07	99.66	99.08	98.97	98.37
	ST59	98.93	99.11	99.10	99.01	99.48	99.05	98.78	98.43
	ST62	99.12	98.75	99.39	98.76	99.23	99.15	98.79	98.10
	ST648	99.65	99.47	99.03	98.91	99.47	99.00	98.73	97.53
G	ST738	99.65	98.84	98.91	99.34	99.83	98.86	99.14	98.52
Average gene conservation		**99.39**	**99.02**	**99.14**	**98.67**	**99.26**	**99.38**	**98.98**	**98.56**

The extent of gene conservation for *acrA, acrB, tolC, eefR, eefA, eefB, eefC,* and *eefD* across phylogroups B2, D, E, F, and G are highlighted as follows ≥96%, ≥97% ≥98% ≥99%

Table 2. Susceptibility of *E. coli* to antimicrobials following loss and gain of *eefABC* and *eefD*

																			MIC (μ
Strain	**AZT**	**BAC**	**CAR**	**CEF**	**CHL**	**CIP**	**CLI**	**ERY**	**EB**	**FA**	**GEN**	**MER**	**MOX**	**NA**	**NOV**	**RIF**	**SPE**	**TET**	**TIC**
ATCC 25922	0.12	32	16	0.06	4	0.03	128	64	512	512	0.5	0.016	0.03	2	32	>32	8	1	8
ATCC 25922 ΔeefB	0.12	32	16	0.06	4	0.016	128	64	512	512	0.5	0.016	0.03	2	32	>32	8	1	8
ATCC 25922 ΔeefABC	0.25	32	16	0.06	2	0.03	128	64	512	512	0.5	0.016	0.03	2	32	>32	8	1	8
ATCC 25922 ΔeefD	0.12	32	16	0.06	8	0.06	256	64	512	1024	0.5	0.016	0.06	8	64	>32	8	1	8
MG1655 + pET21a	0.12	64	>1024	0.06	8	0.016	128	32	1024	512	0.12	0.03	0.06	8	128	>32	4	1	>1024
MG1655 + pET21a *eefABC*	0.12	64	>1024	0.06	8	0.016	128	32	1024	1024	0.25	0.03	0.06	8	128	>32	4	1	>1024
MG1655 + pET24a	0.06	64	8	0.06	8	0.016	128	32	1024	512	0.25	0.03	0.03	8	128	>32	8	1	4
MG1655 + pET24a *eefABC*	0.06	64	8	0.06	8	0.016	128	32	1024	512	0.25	0.016	0.06	8	128	>32	4	1	4
MG1655 + pACYC177	0.06	64	>1024	0.06	8	0.016	128	32	1024	1024	0.5	0.03	0.06	8	128	>32	4	1	>1024
MG1655 + pACYC177 *eefD*	0.06	64	8	0.06	8	0.016	128	32	1024	1024	0.5	0.03	0.06	8	128	>32	4	1	4
MG1655 ΔacrB + pET21a	0.12	4	>1024	<0.016	1	0.008	2	2	8	8	0.12	0.03	<0.008	2	2	32	4	0.25	>1024
MG1655 ΔacrB + pET21a *eefABC*	0.06	8	>1024	0.03	1	0.008	2	2	**32**	16	0.12	0.03	<0.008	2	4	32	4	0.25	>1024
MG1655 ΔacrB + pET24a	0.12	4	8	<0.016	1	0.008	2	2	8	8	0.25	0.03	<0.008	2	2	32	4	0.25	4
MG1655 ΔacrB + pET24a *eefABC*	0.06	8	8	<0.016	2	0.008	2	4	**32**	16	0.25	0.03	<0.008	2	4	32	4	0.25	4
MG1655 ΔacrB + pACYC177	0.12	4	>1024	<0.016	1	0.008	2	2	8	8	0.5	0.03	<0.008	2	2	32	4	0.25	>1024
MG1655 ΔacrB + pACYC177 *eefD*	0.12	4	8	<0.016	1	0.008	2	2	8	8	0.5	0.03	<0.008	2	2	32	4	0.25	4
MG1655 ΔacrB + pET21a eefABC pACYC177 *eefD*					2	0.008													

AZT – aztreonam, BAC – benzalkonium chloride, CAR – carbenicillin, CEF – cefotaxime, CHL – chloramphenicol, CIP – ciprofloxacin, CLI – clindamycin, ERY – erythromycin, EB – ethidium bromide, FA -fusidic acid, GEN – gentamicin, MER – meropenem, MOX – moxifloxacin, NA – nalidixic acid, NOV – novobiocin, RIF – rifampicin, SPE – spectinomycin, TET – tetracycline, TIC – ticarcillin

References

ADAMS, P. P., BANIULYTE, G., ESNAULT, C., CHEGIREDDY, K., SINGH, N., MONGE, M., DALE, R. K., STORZ, G. & WADE, J. T. 2021. Regulatory roles of Escherichia coli 5' UTR and ORF-internal RNAs detected by 3' end mapping. *Elife,* 10.

ADLER, J. & BIBI, E. 2002. Membrane topology of the multidrug transporter MdfA: Complementary gene fusion studies reveal a nonessential C-terminal domain. *Journal of Bacteriology,* 184**,** 3313-3320.

ALAV, I., KOBYLKA, J., KUTH, M. S., POS, K. M., PICARD, M., BLAIR, J. M. A. & BAVRO, V. N. 2021. Structure, Assembly, and Function of Tripartite Efflux and Type 1 Secretion Systems in Gram-Negative Bacteria. *Chem Rev,* 121**,** 5479-5596.

ALCALDE-RICO, M., HERNANDO-AMADO, S., BLANCO, P. & MARTÍNEZ, J. L. 2016. Multidrug Efflux Pumps at the Crossroad between Antibiotic Resistance and Bacterial Virulence. *Frontiers in microbiology*.

ANDERSEN, C., KORONAKIS, E., BOKMA, E., ESWARAN, J., HUMPHREYS, D., HUGHES, C. & KORONAKIS, V. 2002a. Transition to the open state of the TolC periplasmic tunnel entrance. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 11103-11108.

ANDERSEN, C., KORONAKIS, E., HUGHES, C. & KORONAKIS, V. 2002b. An aspartate ring at the TolC tunnel entrance determines ion selectivity and presents a target for blocking by large cations. *Molecular Microbiology,* 44**,** 1131-1139.

ANES, J., MCCUSKER, M. P., FANNING, S. & MARTINS, M. 2015. The ins and outs of RND efflux pumps in Escherichia coli. *Front Microbiol,* 6**,** 587.

BAUGH, S., PHILLIPS, C. R., EKANAYAKA, A. S., PIDDOCK, L. J. & WEBBER, M. A. 2014. Inhibition of multidrug efflux as a strategy to prevent biofilm formation. *J Antimicrob Chemother,* 69**,** 673-81.

BAVRO, V. N., PIETRAS, Z., FURNHAM, N., PEREZ-CANO, L., FERNANDEZ-RECIO, J., PEI, X. Y., MISRA, R. & LUISI, B. 2008. Assembly and channel opening in a bacterial drug efflux machine. *Mol Cell,* 30**,** 114-21.

BAZZINI, S., UDINE, C., SASS, A., PASCA, M. R., LONGO, F., EMILIANI, G., FONDI, M., PERRIN, E., DECOROSI, F., VITI, C., GIOVANNETTI, L., LEONI, L., FANI, R., RICCARDI, G., MAHENTHIRALINGAM, E. & BURONI, S. 2011. Deciphering the role of RND efflux transporters in Burkholderia cenocepacia. *PLoS One,* 6**,** e18902.

BERGLER, H., WALLNER, P., EBELING, A., LEITINGER, B., FUCHSBICHLER, S., ASCHAUER, H., KOLLENZ, G., HOGENAUER, G. & TURNOWSKY, F. 1994. Protein EnvM is the NADH-dependent enoyl-ACP reductase (FabI) of Escherichia coli. *J Biol Chem,* 269**,** 5493-6.

BLAIR, J. M., SMITH, H. E., RICCI, V., LAWLER, A. J., THOMPSON, L. J. & PIDDOCK, L. J. 2015. Expression of homologous RND efflux pump genes is dependent upon AcrB expression: implications for efflux and virulence inhibitor design. *J Antimicrob Chemother,* 70**,** 424-31.

BRENNER, D. J., FANNING, G. R., STEIGERWALT, A. G., ORSKOV, I. & ORSKOV, F. 1972. Polynucleotide sequence relatedness among three groups of pathogenic Escherichia coli strains. *Infect Immun,* 6**,** 308-15.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

CHAN, Y. Y., BIAN, H. S., TAN, T. M., MATTMANN, M. E., GESKE, G. D., IGARASHI, J., HATANO, T., SUGA, H., BLACKWELL, H. E. & CHUA, K. L. 2007. Control of quorum sensing by a Burkholderia pseudomallei multidrug efflux pump. *J Bacteriol,* 189**,** 4320-4.

CHEN, J., KURODA, T., HUDA, M. N., MIZUSHIMA, T. & TSUCHIYA, T. 2003. An RND-type multidrug efflux pump SdeXY from Serratia marcescens. *J Antimicrob Chemother,* 52**,** 176-9.

CHETRI, S., BHOWMIK, D., PAUL, D., PANDEY, P., CHANDA, D. D., CHAKRAVARTY, A., BORA, D. & BHATTACHARJEE, A. 2019. AcrAB-TolC efflux pump system plays a role in carbapenem non-susceptibility in Escherichia coli. *BMC Microbiol,* 19**,** 210.

CLERMONT, O., BONACORSI, S. & BINGEN, E. 2000. Rapid and simple determination of the Escherichia coli phylogenetic group. *Appl Environ Microbiol,* 66**,** 4555-8.

CLSI 2020. *Performance standards for antimicrobial susceptibility testing, 30th edition*, Clinical and Laboratory Standards Institute.

COLCLOUGH, A. L., SCADDEN, J. & BLAIR, J. M. A. 2019. TetR-family transcription factors in Gram-negative bacteria: Conservation, variation and implications for efflux-mediated antimicrobial resistance. *BMC Genomics,* 20.

COUDEYRAS, S., NAKUSI, L., CHARBONNEL, N. & FORESTIER, C. 2008. A tripartite efflux pump involved in gastrointestinal colonization by Klebsiella pneumoniae confers a tolerance response to inorganic acid. *Infection and Immunity*.

DARBY, E. M., BAVRO, V. N., DUNN, S., MCNALLY, A. & BLAIR, J. M. A. 2023. RND pumps across the genus Acinetobacter: AdeIJK is the universal efflux pump. *Microb Genom,* 9.

DELMAR, J. A., SU, C. C. & YU, E. W. 2013. Structural mechanisms of heavy-metal extrusion by the Cus efflux system. *Biometals,* 26**,** 593-607.

DENAMUR, E., CLERMONT, O., BONACORSI, S. & GORDON, D. 2021. The population genetics of pathogenic Escherichia coli. *Nat Rev Microbiol,* 19**,** 37-54.

DOMÍNGUEZ-MEDINA, C. C., PÉREZ-TOLEDO, M., SCHAGER, A. E., MARSHALL, J. L., COOK, C. N., BOBAT, S., HWANG, H., CHUN, B. J., LOGAN, E., BRYANT, J. A., CHANNELL, W. M., MORRIS, F. C., JOSSI, S. E., ALSHAYEA, A., ROSSITER, A. E., BARROW, P. A., HORSNELL, W. G., MACLENNAN, C. A., HENDERSON, I. R., LAKEY, J. H., GUMBART, J. C., LÓPEZ-MACÍAS, C., BAVRO, V. N. & CUNNINGHAM, A. F. 2020. Outer membrane protein size and LPS O-antigen define protective antibody targeting to the Salmonella surface. *Nature Communications,* 11.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

EUCAST 2021. *Routine and extended internal quality control for MIC determination and disk diffusion as recommended by EUCAST. Version 11*, European Society of Clinical Microbiology and Infectious Diseases.

FRANKE, S., GRASS, G. & NIES, D. H. 2001. The product of the ybdE gene of the Escherichia coli chromosome is involved in detoxification of silver ions. *Microbiology,* 147**,** 965-72.

FRICKE, W. F., WRIGHT, M. S., LINDELL, A. H., HARKINS, D. M., BAKER-AUSTIN, C., RAVEL, J. & STEPANAUSKAS, R. 2008. Insights into the environmental

resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate Escherichia coli SMS-3-5. *Journal of Bacteriology*.

GORDON, D. M. & COWLING, A. 2003. The distribution and genetic structure of Escherichia coli in Australian vertebrates: Host and geographic effects. *Microbiology,* 149**,** 3575-3586.

GRASS, G. & RENSING, C. 2001. Genes involved in copper homeostasis in Escherichia coli. *J Bacteriol,* 183**,** 2145-7.

GUAN, L. & NAKAE, T. 2001. Identification of essential charged residues in transmembrane segments of the multidrug transporter MexB of Pseudomonas aeruginosa. *Journal of Bacteriology,* 183**,** 1734-1739.

GUPTA, A., MATSUI, K., LO, J. F. & SILVER, S. 1999. Molecular basis for resistance to silver cations in Salmonella. *Nat Med,* 5**,** 183-8.

HAWKEY, J., MONK, J. M., BILLMAN-JACOBE, H., PALSSON, B. & HOLT, K. E. 2020. Impact of insertion sequences on convergent evolution of Shigella species. *PLOS Genetics,* 16**,** e1008931-e1008931.

HENDERSON, P. J. F., MAHER, C., ELBOURNE, L. D. H., EIJKELKAMP, B. A., PAULSEN, I. T. & HASSAN, K. A. 2021. Physiological Functions of Bacterial "Multidrug" Efflux Pumps. *Chem Rev,* 121**,** 5417-5478.

HENG, J., ZHAO, Y., LIU, M., LIU, Y., FAN, J., WANG, X., ZHAO, Y. & ZHANG, X. C. 2015. Substrate-bound structure of the E. coli multidrug resistance transporter MdfA. *Cell Res,* 25**,** 1060-73.

HIGGINS, M. K., ESWARAN, J., EDWARDS, P., SCHERTLER, G. F., HUGHES, C. & KORONAKIS, V. 2004. Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug efflux protein TolC. *J Mol Biol,* 342**,** 697-702.

HOLMES, C. L., ANDERSON, M. T., MOBLEY, H. L. T. & BACHMAN, M. A. 2021. Pathogenesis of Gram-Negative Bacteremia. *Clin Microbiol Rev,* 34.

IMUTA, N., NISHI, J., TOKUDA, K., FUJIYAMA, R., MANAGO, K., IWASHITA, M., SARANTUYA, J. & KAWANO, Y. 2008. The Escherichia coli efflux pump TolC promotes aggregation of enteroaggregative E. coli 042. *Infection and Immunity,* 76**,** 1247-1256.

KATOH, K., ROZEWICKI, J. & YAMADA, K. D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform,* 20**,** 1160-1166.

KIM, H. M., XU, Y., LEE, M., PIAO, S., SIM, S. H., HA, N. C. & LEE, K. 2010. Functional relationships between the AcrA hairpin tip region and the TolC aperture tip region for the formation of the bacterial tripartite efflux pump AcrAB-TolC. *Journal of Bacteriology,* 192**,** 4498-4503.

KIM, J., WEBB, A. M., KERSHNER, J. P., BLASKOWSKI, S. & COPLEY, S. D. 2014. A versatile and highly efficient method for scarless genome editing in Escherichia coli and Salmonella enterica. *BMC Biotechnology 2014 14:1,* 14**,** 1-13.

KLENOTIC, P. A., MOSENG, M. A., MORGAN, C. E. & YU, E. W. 2020. Structural and Functional Diversity of Resistance–Nodulation–Cell Division Transporters. *Chemical Reviews***,** acs.chemrev.0c00621-acs.chemrev.0c00621.

KORONAKIS, V., SHARFF, A., KORONAKIS, E., LUISI, B. & HUGHES, C. 2000. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature,* 405**,** 914-9.

KULATHILA, R., KULATHILA, R., INDIC, M. & VAN DEN BERG, B. 2011. Crystal structure of Escherichia coli CusC, the outer membrane component of a heavy metal efflux pump. *PLoS One,* 6**,** e15610.

LECOINTRE, G., RACHDI, L., DARLU, P. & DENAMUR, E. 1998. Escherichia coli Molecular Phylogeny Using the Incongruence Length Difference Test. *Mol. Biol. Evol,* 15**,** 1685-1695.

LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Web Server issue Published online,* 47.

LEUS, I. V., ADAMIAK, J., TRINH, A. N., SMITH, R. D., SMITH, L., RICHARDSON, S., ERNST, R. K. & ZGURSKAYA, H. I. 2020. Inactivation of AdeABC and AdeIJK efflux pumps elicits specific nonoverlapping transcriptional and phenotypic responses in Acinetobacter baumannii. *Mol Microbiol,* 114**,** 1049-1065.

LEWINSON, O., ADLER, J., POELARENDS, G. J., MAZURKIEWICZ, P., DRIESSEN, A. J. M. & BIBI, E. 2003. The Escherichia coli multidrug transporter MdfA catalyzes both electrogenic and electroneutral transport reactions. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 1667-1672.

LI, D., ELANKUMARAN, P., KUDINHA, T., KIDSLEY, A. K., TROTT, D. J., JAROCKI, V. M. & DJORDJEVIC, S. P. 2023. Dominance of Escherichia coli sequence types ST73, ST95, ST127 and ST131 in Australian urine isolates: a genomic analysis of antimicrobial resistance and virulence linked to F plasmids. *Microb Genom,* 9.

LONG, F., SU, C. C., LEI, H. T., BOLLA, J. R., DO, S. V. & YU, E. W. 2012. Structure and mechanism of the tripartite CusCBA heavy-metal efflux complex. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1047-58.

MA, K. C., MORTIMER, T. D., HICKS, A. L., WHEELER, N. E., SANCHEZ-BUSO, L., GOLPARIAN, D., TAIAROA, G., RUBIN, D. H. F., WANG, Y., WILLIAMSON, D. A., UNEMO, M., HARRIS, S. R. & GRAD, Y. H. 2020. Adaptation to the cervical environment is associated with increased antibiotic susceptibility in Neisseria gonorrhoeae. *Nat Commun,* 11**,** 4126.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A,* 95**,** 3140-5.

MARSHALL, R. L. & BAVRO, V. N. 2020. Mutations in the TolC Periplasmic Domain Affect Substrate Specificity of the AcrAB-TolC Pump. *Front Mol Biosci,* 7**,** 166.

MASI, M., PAGES, J. M. & PRADEL, E. 2006. Production of the cryptic EefABC efflux pump in Enterobacter aerogenes chloramphenicol-resistant mutants. *J Antimicrob Chemother,* 57**,** 1223-6.

MASI, M., PAGES, J. M., VILLARD, C. & PRADEL, E. 2005. The eefABC multidrug efflux pump operon is repressed by H-NS in Enterobacter aerogenes. *J Bacteriol,* 187**,** 3894-7.

MASUDA, N., SAKAGAWA, E., OHYA, S., GOTOH, N., TSUJIMOTO, H. & NISHINO, T. 2000. Substrate specificities of MexAB-OprM, MexCD-OprJ, and MexXY-OprM efflux pumps in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy,* 44**,** 3322-3327.

MATEUS, C., NUNES, A. R., OLEASTRO, M., DOMINGUES, F. & FERREIRA, S. 2021. RND Efflux Systems Contribute to Resistance and Virulence of Aliarcobacter butzleri. *Antibiotics (Basel),* 10.

MCNEIL, H. E., ALAV, I., TORRES, R. C., ROSSITER, A. E., LAYCOCK, E., LEGOOD, S., KAUR, I., DAVIES, M., WAND, M., WEBBER, M. A., BAVRO, V.

N. & BLAIR, J. M. A. 2019. Identification of binding residues between periplasmic adapter protein (PAP) and RND efflux pumps explains PAP-pump promiscuity and roles in antimicrobial resistance. *PLoS Pathog,* 15**,** e1008101.

MONLEZUN, L., PHAN, G., BENABDELHAK, H., LASCOMBE, M. B., ENGUÉNÉ, V. Y. N., PICARD, M. & BROUTIN, I. 2015. New OprM structure highlighting the nature of the N-terminal anchor. *Frontiers in Microbiology,* 6.

MORAN, N. A. & PLAGUE, G. R. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev,* 14**,** 627-33.

NAGARATHINAM, K., NAKADA-NAKURA, Y., PARTHIER, C., TERADA, T., JUGE, N., JAENECKE, F., LIU, K., HOTTA, Y., MIYAJI, T., OMOTE, H., IWATA, S., NOMURA, N., STUBBS, M. T. & TANABE, M. 2018. Outward open conformation of a Major Facilitator Superfamily multidrug/H+ antiporter provides insights into switching mechanism. *Nature Communications,* 9.

NAKASHIMA, R., SAKURAI, K., YAMASAKI, S., HAYASHI, K., NAGATA, C., HOSHINO, K., ONODERA, Y., NISHINO, K. & YAMAGUCHI, A. 2013. Structural basis for the inhibition of bacterial multidrug exporters. *Nature,* 500**,** 102-106.

NEMEC, A., MAIXNEROVA, M., VAN DER REIJDEN, T. J., VAN DEN BROEK, P. J. & DIJKSHOORN, L. 2007. Relationship between the AdeABC efflux system gene content, netilmicin susceptibility and multidrug resistance in a genotypically diverse collection of Acinetobacter baumannii strains. *J Antimicrob Chemother,* 60**,** 483-9.

NI, R. T., ONISHI, M., MIZUSAWA, M., KITAGAWA, R., KISHINO, T., MATSUBARA, F., TSUCHIYA, T., KURODA, T. & OGAWA, W. 2020. The role of RND-type efflux pumps in multidrug-resistant mutants of Klebsiella pneumoniae. *Sci Rep,* 10**,** 10876.

NISHINO, K., LATIFI, T. & GROISMAN, E. A. 2006. Virulence and drug resistance roles of multidrug efflux systems of *Salmonella enterica* serovar Typhimurium. *Mol Microbiol,* 59**,** 126-41.

NISHINO, K. & YAMAGUCHI, A. 2001. Analysis of a complete library of putative drug transporter genes in Escherichia coli. *J Bacteriol,* 183**,** 5803-12.

NOWAK, J., SEIFERT, H. & HIGGINS, P. G. 2015. Prevalence of eight resistance-nodulation-division efflux pump genes in epidemiologically characterized Acinetobacter baumannii of worldwide origin. *J Med Microbiol,* 64**,** 630-635.

NTSOGO ENGUÉNÉ, Y. V., PHAN, G., GARNIER, C., DUCRUIX, A., PRANGÉ, T. & BROUTIN, I. 2017. Xenon for tunnelling analysis of the efflux pump component OprN. *PLoS ONE,* 12.

ONDOV, B. D., TREANGEN, T. J., MELSTED, P., MALLONEE, A. B., BERGMAN, N. H., KOREN, S. & PHILLIPPY, A. M. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol,* 17**,** 132.

PADILLA, E., LLOBET, E., DOMENECH-SANCHEZ, A., MARTINEZ-MARTINEZ, L., BENGOECHEA, J. A. & ALBERTI, S. 2010. Klebsiella pneumoniae AcrAB efflux pump contributes to antimicrobial resistance and virulence. *Antimicrob Agents Chemother,* 54**,** 177-83.

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T., FOOKES, M., FALUSH, D., KEANE, J. A. & PARKHILL, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3.

PAK, J. E., EKENDE, E. N., KIFLE, E. G., O'CONNELL, J. D., 3RD, DE ANGELIS, F., TESSEMA, M. B., DERFOUFI, K. M., ROBLES-COLMENARES, Y., ROBBINS, R. A., GOORMAGHTIGH, E., VANDENBUSSCHE, G. & STROUD, R. M. 2013.

Structures of intermediate transport states of ZneA, a Zn(II)/proton antiporter. *Proc Natl Acad Sci U S A,* 110**,** 18484-9.

PUGH, H. L., CONNOR, C., SIASAT, P., MCNALLY, A. & BLAIR, J. M. A. 2023. E. coli ST11 (O157:H7) does not encode a functional AcrF efflux pump. *Microbiology (Reading),* 169.

QUISTGAARD, E. M., LOW, C., GUETTOU, F. & NORDLUND, P. 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat Rev Mol Cell Biol,* 17**,** 123-32.

REDDY, V. S., SHLYKOV, M. A., CASTILLO, R., SUN, E. I. & SAIER, M. H. 2012. The major facilitator superfamily (MFS) revisited. *FEBS Journal,* 279**,** 2022-2035.

ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research,* 42.

ROBINSON, O., DYLUS, D. & DESSIMOZ, C. 2016. Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. *Molecular Biology and Evolution,* 33**,** 2163-2166.

SANCHEZ, L., PAN, W., VINAS, M. & NIKAIDO, H. 1997. The acrAB homolog of Haemophilus influenzae codes for a functional multidrug efflux pump. *Journal of Bacteriology*.

SCHULZ, R. & KLEINEKATHÖFER, U. 2009. Transitions between closed and open conformations of TolC: The effects of ions in simulations. *Biophysical Journal,* 96**,** 3116-3125.

SEEGER, M. A., SCHIEFNER, A., EICHER, T., VERREY, F., DIEDERICHS, K. & POS, K. M. 2006. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science,* 313**,** 1295-1298.

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SEEMANN, T. 2017. ABRicate: Mass screening of contigs for antimicrobial resistance or virulence genes. *Github*.

SHAIK, S., RANJAN, A., TIWARI, S. K., HUSSAIN, A., NANDANWAR, N., KUMAR, N., JADHAV, S., SEMMLER, T., BADDAM, R., ISLAM, M. A., ALAM, M., WIELER, L. H., WATANABE, H. & AHMED, N. 2017. Comparative Genomic Analysis of Globally Dominant ST131 Clone with Other Epidemiologically Successful Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. *mBio,* 8.

SMITH, H. E. & BLAIR, J. M. 2014. Redundancy in the periplasmic adaptor proteins AcrA and AcrE provides resilience and an ability to export substrates of multidrug efflux. *J Antimicrob Chemother,* 69**,** 982-7.

SONG, S., HWANG, S., LEE, S., HA, N. C. & LEE, K. 2014. Interaction mediated by the putative tip regions of MdsA and MdsC in the formation of a Salmonella-specific tripartite efflux pump. *PLoS ONE,* 9.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *BIOINFORMATICS APPLICATIONS,* 30**,** 1312-1313.

SUGIYAMA, Y., NAKAMURA, A., MATSUMOTO, M., KANBE, A., SAKANAKA, M., HIGASHI, K., IGARASHI, K., KATAYAMA, T., SUZUKI, H. & KURIHARA, S. 2016. A Novel Putrescine Exporter SapBCDF of Escherichia coli. *Journal of Biological Chemistry,* 291**,** 26343-26351.

SULLIVAN, M. J., PETTY, N. K. & BEATSON, S. A. 2011. Easyfig: A genome comparison visualizer. *Bioinformatics,* 27**,** 1009-1010.

SWICK, M. C., MORGAN-LINNELL, S. K., CARLSON, K. M. & ZECHIEDRICH, L. 2011. Expression of multidrug efflux pump genes acrAB-tolC, mdfA, and norE in Escherichia coli clinical isolates as a function of fluoroquinolone and multidrug resistance. *Antimicrob Agents Chemother,* 55**,** 921-4.

SYMMONS, M. F., BOKMA, E., KORONAKIS, E., HUGHES, C. & KORONAKIS, V. 2009. The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proc Natl Acad Sci U S A,* 106**,** 7173-8.

TENAILLON, O., SKURNIK, D., PICARD, B. & DENAMUR, E. 2010. The population genetics of commensal Escherichia coli. *Nat Rev Microbiol,* 8**,** 207-17.

THOMASON, M. K., BISCHLER, T., EISENBART, S. K., FORSTNER, K. U., ZHANG, A., HERBIG, A., NIESELT, K., SHARMA, C. M. & STORZ, G. 2015. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. *J Bacteriol,* 197**,** 18-28.

TOUCHON, M. & ROCHA, E. P. 2007. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol,* 24**,** 969-81.

VACCARO, L., SCOTT, K. A. & SANSOM, M. S. P. 2008. Gating at both ends and breathing in the middle: Conformational dynamics of TolC. *Biophysical Journal,* 95**,** 5681-5691.

VAN DEN BELD, M. J. & REUBSAET, F. A. 2012. Differentiation between Shigella, enteroinvasive Escherichia coli (EIEC) and noninvasive Escherichia coli. *Eur J Clin Microbiol Infect Dis,* 31**,** 899-904.

WANG, Z., FAN, G., HRYC, C. F., BLAZA, J. N., SERYSHEVA, I. I., SCHMID, M. F., CHIU, W., LUISI, B. F. & DU, D. 2017. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *eLife,* 6.

WANG-KAN, X., BLAIR, J. M. A., CHIRULLO, B., BETTS, J., LA RAGIONE, R. M., IVENS, A., RICCI, V., OPPERMAN, T. J. & PIDDOCK, L. J. V. 2017. Lack of AcrB Efflux Function Confers Loss of Virulence on Salmonella enterica Serovar Typhimurium. *mBio,* 8.

WANG-KAN, X., RODRIGUEZ-BLANCO, G., SOUTHAM, A. D., WINDER, C. L., DUNN, W. B., IVENS, A. & PIDDOCK, L. J. V. 2021. Metabolomics Reveal Potential Natural Substrates of AcrB in Escherichia coli and Salmonella enterica Serovar Typhimurium. *mBio,* 12.

WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research,* 46**,** W296-W303.

WEAVER, J., MOHAMMAD, F., BUSKIRK, A. R. & STORZ, G. 2019. Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes. *mBio,* 10.

WIECZOREK, P., SACHA, P., CZABAN, S., HAUSCHILD, T., OJDANA, D., KOWALCZUK, O., MILEWSKI, R., PONIATOWSKI, B., NIKLINSKI, J. & TRYNISZEWSKA, E. 2013. Distribution of AdeABC efflux system genes in genotypically diverse strains of clinical Acinetobacter baumannii. *Diagn Microbiol Infect Dis,* 77**,** 106-9.

WIRTH, T., FALUSH, D., LAN, R., COLLES, F., MENSA, P., WIELER, L. H., KARCH, H., REEVES, P. R., MAIDEN, M. C. J., OCHMAN, H. & ACHTMAN, M. 2006. Sex and virulence in Escherichia coli: An evolutionary perspective. *Molecular Microbiology,* 60**,** 1136-1151.

XIE, Z. & TANG, H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics,* 33**,** 3340-3347.

YANG, F., YANG, J., ZHANG, X., CHEN, L., JIANG, Y., YAN, Y., TANG, X., WANG, J., XIONG, Z., DONG, J., XUE, Y., ZHU, Y., XU, X., SUN, L., CHEN, S., NIE, H., PENG, J., XU, J., WANG, Y., YUAN, Z., WEN, Y., YAO, Z., SHEN, Y., QIANG, B., HOU, Y., YU, J. & JIN, Q. 2005. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res,* 33**,** 6445-58.

YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods,* 12**,** 7-8.

YANG, L., CHEN, L., SHEN, L., SURETTE, M. & DUAN, K. 2011. Inactivation of MuxABC-OpmB transporter system in Pseudomonas aeruginosa leads to increased ampicillin and carbenicillin resistance and decreased virulence. *J Microbiol,* 49**,** 107-14.

YIN, Y., HE, X., SZEWCZYK, P., NGUYEN, T. & CHANG, G. 2006. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science,* 312**,** 741-744.

YONEHARA, R., YAMASHITA, E. & NAKAGAWA, A. 2016. Crystal structures of OprN and OprJ, outer membrane factors of multidrug tripartite efflux pumps of Pseudomonas aeruginosa. *Proteins: Structure, Function and Bioinformatics,* 84**,** 759-769.

ZHOU, Z., ALIKHAN, N. F., MOHAMED, K., FAN, Y., AGAMA STUDY, G. & ACHTMAN, M. 2020. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. *Genome Res,* 30**,** 138-152.

Acknowledgements

We thank Karl Hassan for his insightful comments and suggestions in the viva of Hannah Pugh which helped shape sections of this work.

This work was supported by the BBSRC MiBTP DTP grant awarded to HLP (BB/M01116X/1). EMD was funded by the Wellcome Trust (222386/Z/21/Z).

Competing interest statement

None to declare