Diversity and conservation of RND efflux pumps across *Escherichia coli*

by

Hannah Louise Pugh

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection

College of Medical and Dental Sciences

University of Birmingham

November 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

RND efflux pumps are located in the inner membrane of Gram-negative bacteria and form tripartite complexes that span the inner membrane, periplasm and outer membrane. It is widely accepted that *Escherichia coli* isolates possess six RND pumps; AcrB, AcrD, AcrF, MdtBC, MdtF and CusA. These pumps are involved in a wide range of physiological functions but are best characterised for their role in antimicrobial efflux. Furthermore, overexpression of RND pumps in clinical isolates confers an MDR phenotype. Though once a clinical isolate, *E. coli* K-12 and derivatives such as K-12 MG1655 have lost virulence factors and contain significant mutations as a result of years of laboratory passage. Despite this most work determining the structure and function of *E. coli* RND pumps has been carried out using *E. coli* K-12 MG1655.

Work in this study found that RND efflux pumps were highly conserved across *E. coli* although level of conservation varied between systems. In addition, when using the K-12 protein sequences as a reference, amino acid substitutions were present and highly conserved across 20 *E. coli* lineages. Moreover AcrA, AcrB and TolC which form the main RND efflux pump in *E. coli* each possessed an amino acid substitution which was present in almost 100% of assemblies belonging to phylogroups B2, D and F. These phylogroups are most commonly associated with extraintestinal *E. coli* infections and AMR carriage. Furthermore, ST11 which is commonly referred to as *E. coli* O157:H7 was found to lack to a functional AcrF protein due to a highly conserved 45 nucleotide insertion half way through the gene which encodes two stop codons.

An additional seventh RND protein, EefB, was also identified in the work and found to be highly prevalent across phylogroups B2, D, E, F and G but completely absent from phylogroups A, B1 and C which are more commonly associated with

commensal living. This RND pump was found encoded on a five gene operon that encoded a TetR regulator (EefR), PAP (EefA), RND pump (EefB), OMP (EefC), and an MFS transporter (EefD). In *E. coli* which encode the operon, the genes were found to be highly conserved suggesting strong positive selection for an important biological function.

Taken together, this study demonstrates that RND pumps are highly conserved across *E. coli* but the extent to which systems are conserved and also the number of systems present varies across *E. coli* phylogroups. In addition it shows that the use of a model strain not only removes strain specific diversity but prevents the identification and subsequent characterisation of RND systems that are highly prevalent across a species.

Acknowledgements

Thank you to the BBSRC for funding this project and the MIBTP programme for giving me the opportunity to undertake a PhD.

A huge thanks goes to my supervisor, Dr. Jessica Blair. Your continuous support and friendship have been invaluable. You saved the day when I was ready to walk away and sell my soul to the financial sector. I would also like to thank Prof. Alan McNally and Prof. Willem van Schaik for their guidance and help (and the free pints).

Thank you to the Blair lab group, both past and present for being so much fun and always being up for coffee and cake. Shout out to Ilyas, Lizzy, Enea, and Pauline in particular for the moral support during the last few months.

I would like to thank Ms. A. Whitcombe for sparking a passion for biology and Dr. Beatrix Fahnert for introducing me to the world of microbiology and providing me the opportunities that have enabled me to reach where I am today.

Thanks to Beth and Chris for always being on hand, and for the pre-Covid coffee and cake breaks and the post-Covid beer garden trips. To Rach and Lizzie for physically getting me through the first two years. And to Llinos, Jo, Caitlin and Evan for being just being your awesome crazy selves. I don't know what I'd do without all of you.

I would also like to acknowledge my grandparents who instilled in me the importance of an education very early on. Only my Nana will see my finished thesis but I hope Grandpa Tom, Gran and Grandpa Roy are proud wherever they are.

Finally a huge thanks goes to my parents and sister for their unwavering support from undergraduate through to this moment. You guys have believed in me when I couldn't. And to Janet for allowing science and this document to consume both of our lives for the last four years. Let the social life resume!

Clarification of Results produced in Collaboration

Chapter 4

Escherichia coli pangenomes used to determine RND gene and protein conservation were constructed by Christopher Connor.

Chapter 6

The EefABC protein homology model was produced by Dr. Vassiliy Bavro at the University of Essex.

Table of contents

Chapter 1 Introduction 1

1.1 Genome sequencing and the bacterial pangenome 1

1.2 Bacterial genome plasticity 4

1.3 *Escherichia coli* 5

1.4 *E. coli* infection 5

1.5 *E. coli* typing and classification 9

1.6 Mechanisms of resistance 15

1.7 Multidrug resistance through efflux 20

1.8 RND systems in *E. coli* 25

1.9 AcrAB-TolC 29

1.10 AcrAB-TolC regulation 32

1.11 The wider role of RND efflux pumps 36

1.12 Efflux inhibition 37

1.13 Sequence variation and/or conservation across RND efflux pumps 39

1.14 Aims of work 40

Chapter 2 Materials and methods 42

2.1 Basic microbiology 42

2.2 Isolation of plasmid DNA 49

2.3 Production of electrocompetent cells 49

2.4 Bacterial transformation 50

2.5 PCR amplification conditions 50

2.6 DNA purification 58

2.7 DNA sequencing of PCR products 59

2.8 Gene cloning 59

2.8.1 Cloning of *acrF* from K-12 and ST11 into pET21a 59

2.8.2 Cloning of *eefABC* into pET21a and pET24a 63

2.8.3 Cloning of *eefD* into pACYC177 66

2.9 Gene disruption in *E. coli* 68

2.9.1 Homologous recombination 68

2.9.2 Bacteriophage transduction 70

2.10 Phenotypic characterisation assays 72

2.10.1 Ciprofloxacin accumulation 72

2.10.2 Quantification of biofilm production 73

2.10.3 Measurement of antimicrobial susceptibility by agar dilution method 74

2.10.4 Measurement of curli and cellulose production using Congo red 76

2.10.5 Ethidium bromide efflux 76

2.10.6 1,2'-dinaphthylamine efflux 77

2.11 Statistics 79

2.12 Bioinformatics 79

2.12.1 Chapters 1, 2 and 3 79

2.12.2 Chapter 4 81

Chapter 3 Conservation of RND genes and proteins across 20 *Escherichia coli* sequence types 86

3.1 Background 86

3.2 Aims 87

3.3 Hypotheses 87

3.4 Identification of RND genes across *E. coli* pangenomes 87

3.5 The number of RND gene alleles varies between *E. coli* pangenomes 88

3.6 Diversity in protein sequences at the individual ST and phylogroup level 93

3.6.1 AcrAB-TolC 93

3.6.2 AcrZ 101

3.6.3 AcrD 101

3.6.4 AcrEF 106

3.6.5 CusABC 111

3.6.6 MdtABC 119

3.6.7 MdtEF 124

3.7 Discussion 132

3.8 Key findings 137

3.9 Future work 138

Chapter 4 The *E. coli* lineage ST11 does not possess a functional AcrF protein 139

4.1 Background 139

4.2 Aims 140

4.3 Hypothesis 140

4.4 *acrF* is absent from the ST11 pangenome 140

4.5 *acrF* contains a highly conserved insertion which encodes two internal stop codons 141

4.6 AcrF is non-functional in ST11 144

4.7 Discussion 152

4.8 Key findings 154

4.9 Future work 154

Chapter 5 Characterisation of highly conserved amino acid substitutions present in AcrA, AcrB and TolC 155

5.1 Background 155

5.2 Aims 156

5.3 Hypotheses 156

5.4 The prevalence of specific amino acid substitutions across AcrA, AcrB and TolC in *E. coli* phylogroups 156

5.5 The polymorphisms were not found to alter susceptibility to dyes or antimicrobials. 161

5.6 Intracellular accumulation of ciprofloxacin is not affected by the introduction of the conserved residues 164

5.7 AcrA T104A, AcrB H596N and TolC A233T did not alter the efflux rates of ethidium bromide and 1,2'-dinaphthylamine 168

5.8 Curli and cellulose production are not affected by the introduction of AcrA T104A, AcrB H596N or TolC A233T. 174

5.9 Biofilm production is unaffected by the introduction of the conserved mutant residues 178

5.10 Discussion 181

5.10.1 Structural location of the variants 181

5.10.2 Phylogeny 186

5.10.3 Phenotype 187

5.11 Key findings 188

5.12 Future work 188

Chapter 6 Identification, conservation and characterisation of the novel *eefRABCD* operon across *E. coli* 190

6.1 Background 190

6.2 Aims 190

6.3 Hypotheses 191

6.4 The EefABC RND efflux pump in *E. coli* is not limited to *E. coli* SMS-3-5 191

6.5 The *eefRABCD* operon is highly prevalent across *E. coli* 191

6.6 Despite belonging to phylogroups D and E, ST69 and ST182 do not possess the *eef* operon 200

6.7 *eefRABCD* is conserved across phylogroups B2, D, E, F and G but has been lost in A, B1 and C 200

6.8 Genes that co-occur with the *eefRABCD* operon in *E. coli* 206

6.9 Modelling suggests EefABC is most similar to RND systems from other organisms 214

6.10 Functional characterisation of EefABC 216

6.10.1 Selection of a laboratory strain for *eef* work 216

6.10.2 EefABC is not a drug transport pump 217

6.11 Characterisation of the associated MFS pump, *eefD* 230

6.12 Presence or absence of *eefRABCD* in other Gram negative bacterial species 230

6.13 Discussion 239

6.14 Key findings 243

6.15 Future work 244

Chapter 7 Final Discussion 246

Chapter 8 Appendices 249

8.1 Sequences of RND genes used in Chapters 3, 4 and 5 249

8.2 Sequences of *eef* genes used in Chapter 6 261

8.3 Alignment of *E. coli* SMS-3-5 with *E. coli* ATCC 25922 265

Chapter 9 References 266

List of figures

Figure 1.1 Venn diagram representation of a bacterial pangenome 3

Figure 1.2 Phylogenetic tree of *Escherichia coli* 13

Figure 1.3 Proposed diversification of *Escherichia coli* 14

Figure 1.4 Mechanisms of resistance..... 19

Figure 1.5 Classes of drug efflux pumps present in *E. coli* 24

Figure 1.6 The structure of AcrAB-TolC. 26

Figure 1.7. Functionally rotating mechanism of transport by AcrB 31

Figure 1.8 Regulation of AcrAB-TolC in Enterobacteriaceae 35

Figure 2.1 Plasmid map for pET21 *acrF* K-12 61

Figure 2.2 Plasmid map for pET21a *acrF* ST11 62

Figure 2.3 Plasmid map for pET21a *eefABC* 64

Figure 2.4 Plasmid map for pET24a *eefABC* 65

Figure 2.5 Plasmid map for pACYC177 *eefD*..... 67

Figure 3.1 Conservation of the AcrA protein across 20 *E. coli* STs 95

Figure 3.2 Conservation of the AcrB protein across 20 *E. coli* STs 97

Figure 3.3 Conservation of the TolC protein across 20 *E. coli* STs..... 99

Figure 3.4 Conservation of the AcrZ protein across 20 *E. coli* STs..... 102

Figure 3.5 Conservation of the AcrD protein across 20 *E. coli* STs 104

Figure 3.6 Conservation of the AcrE protein across 20 *E. coli* STs 107

Figure 3.7 Conservation of the AcrF protein across 20 *E. coli* STs..... 109

Figure 3.8 Conservation of the CusA protein across 20 *E. coli* STs 113

Figure 3.9 Conservation of the CusB protein across 20 *E. coli* STs 115

Figure 3.10 Conservation of the CusC protein across 20 *E. coli* STs 117

Figure 3.11 Conservation of the MdtA protein across 20 *E. coli* STs 120
Figure 3.12 Conservation of the MdtB protein across 20 *E. coli* STs 122
Figure 3.13 Conservation of the MdtC protein across 20 *E. coli* STs 126
Figure 3.14 Conservation of the MdtE protein across 20 *E. coli* STs 128
Figure 3.15 Conservation of the MdtF protein across 20 *E. coli* STs 130
Figure 4.1 *acrF* in *E. coli* ST11 has a 45 nucleotide insertion 142
Figure 4.2 Cloning strategy to confirm AcrF in ST11 is non-functional 146
Figure 4.3 Confirmation of *acrF* cloning 147
Figure 4.4 Interruption of the *acrF* gene in K-12 and K-12 Δ*acrB* 149
Figure 5.1 Distribution of AcrA T104A, AcrB H596N and TolC A233T across the *E. coli* assemblies used in the pangenome and RND protein conservation analyses 159
Figure 5.2 Ciprofloxacin accumulation of *E. coli* K-12, Δ*acrA*, Δ*acrB* and Δ*tolC* and strains complemented with both K-12 and amino acid variants 166
Figure 5.3 Efflux of ethidium bromide in *E. coli* K-12, Δ*tolC* and complemented strains 169
Figure 5.4 Efflux of 1,2'-dinaphthylamine by *E. coli* K-12, Δ*acrA*, Δ*acrB* and Δ*tolC* and complemented strains 172
Figure 5.5 Curli and cellulose biogenesis of *E. coli* K-12 MG1655, Δ*acrA*, Δ*acrB*, Δ*tolC* and complemented strains 176
Figure 5.6 Biofilm production in in *E. coli* K-12 MG1655, Δ*acrA*, Δ*acrB*, Δ*tolC* and strains complemented with K-12 and amino acid variants 179
Figure 5.7 Position of T104 within the AcrA monomer 183
Figure 5.8 Position of H596 within in the AcrB monomer and homotrimer 184
Figure 5.9 Position of A233 within the TolC monomer and trimer 185

Figure 6.1 Phylogenetic structure of the assemblies used in this analysis and the distribution of *eefRABCD* 198
Figure 6.2 The genomic region of strains with and without *eefRABCD* 203
Figure 6.3 Representative homology to *eef* operon in *E. coli* STs that do not possess the operon 205
Figure 6.4 Gene network of clusters that co-occur with *eefR* 209
Figure 6.5 Gene network of clusters that co-occur with *eefA* 210
Figure 6.6 Gene network of clusters that co-occur with *eefB* 211
Figure 6.7 Gene network of clusters that co-occur with *eefC* 212
Figure 6.8 Gene network of clusters that co-occur with *eefD* 213
Figure 6.9 Predicted structure of EefABC compared to AcrAB-TolC 215
Figure 6.10 Confirmation of pET21a *eefABC* cloning 218
Figure 6.11 Confirmation of pET24a *eefABC* cloning 220
Figure 6.12 Approach used to inactivate *eefABC* 223
Figure 6.13 Confirmation of *eefB* and *eefABC* gene inactivation 224
Figure 6.14 Alignment of the *E. coli* and *Klebsiella* spp. *eef* operons 231
Figure 6.15 Alignment of the *eef* operon of multiple *Escherichia* spp. 232
Figure 6.16 Operon comparison between *E. coli* and *S. dysenteriae* 238

List of tables

Table 2.1 Strains used in the study 43

Table 2.2 Plasmids used in this study 48

Table 2.3 Primers used in this study 52

Table 2.4 Generic PCR conditions 58

Table 2.5 Compounds used in MIC 75

Table 3.1 Pangenome alleles of RND-associated genes across 20 *E. coli* sequence types 92

Table 4.1 Susceptibility of K-12 Δ*acrB* Δ*acrF* following complementation with *acrF* from K-12 and ST11 151

Table 5.1 Prevalence of AcrA T104A, AcrB H596N and TolC A233T across 20 *E. coli* ST 158

Table 5.2 Prevalence of AcrA T104A, AcrB H596N and TolC A233T across the *E. coli* phylogroups 158

Table 5.3 Antimicrobial susceptibility of strains complemented with K-12 or variant versions of for *acrA*, *acrB* and *tolC* 163

Table 5.4 Rate of ethidium bromide efflux in K-12, K-12 Δ*tolC*, K-12 Δ*tolC* pET21a K-12 tolC and K-12 Δ*tolC* pET21a *tolC* A233T 170

Table 5.5 Rate of 1,2'-diphnaphthylamine efflux in *E. coli* K-12 Δ*acrA*, Δ*acrB* and Δ*tolC*, and complemented strains 173

Table 6.1 The *eefRABCD* operon and its proposed products 194

Table 6.2 *E. coli* assemblies used in this analysis 195

Table 6.3 Presence and subsequent gene coverage of *eefRABCD* across the *E. coli* assemblies with reference to the *E. coli* ATCC 25922 operon 196

Table 6.4 Conservation as percentage nucleotide identity of *eefRABCD* with reference to *E. coli* ATCC 25922 across phylogroups of *E. coli* found to possess the operon 197

Table 6.5 Genes found to co-occur with the *eef* operon and their respective function 208

Table 6.6 Susceptibility of *E. coli* to antimicrobials following loss and gain of *eefABC* and *eefD*..... 228

Table 6.7 Susceptibility of *E. coli* to dyes following loss and gain of *eefABC* and *eefD* 229

Table 6.8 *Shigella* assemblies used in this analysis 234

Table 6.9 Presence and subsequent gene coverage of *eefRABCD* across the *Shigella* species with reference to the *E. coli* ATCC 25922 operon..... 234

Table 6.10 Conservation as percentage nucleotide identity of *eefRABCD* with reference to *E. coli* ATCC 25922 across *Shigella* species..... 236

Abbreviations

1,2'-DNA	-	1,2'-dinaphthylamine
ABC	-	ATP-binding cassette transporter
AIEC	-	adherent invasive *Escherichia coli*
AMR	-	antimicrobial resistance
BLASTN	-	basic local alignment search tool nucleotide
BLASTP	-	basic local alignment search tool protein
bp	-	base pairs
CCCP	-	carbonyl cyanide m-chlorophenyl hydrazone
CDS	-	coding sequence
CFU	-	colony forming units
DAEC	-	diffuse-adhering *Escherichia coli*
DNA	-	deoxyribonucleic acid
EAEC	-	enteroaggregative *Escherichia coli*
ECOR	-	*Escherichia coli* reference collection
EHEC	-	enterohaemorrhagic *Escherichia coli*
EIEC	-	enteroinvasive *Escherichia coli*
EPEC	-	enteropathogenic *Escherichia coli*
ESBL	-	extended spectrum β-lactamase
ETEC	-	enterotoxigenic *Escherichia coli*
ExPEC	-	extraintestinal *Escherichia coli*
HGT	-	horizontal gene transfer
HTH	-	helix-turn-helix
HUS	-	haemolytic uremic syndrome

InPEC	-	intestinal pathogenic *Escherichia coli*
IS	-	insertion sequence
KPC	-	*Klebsiella pneumoniae* carbapenemases
MATE	-	multidrug and toxic compound extrusion
mcr-1	-	mobile colistin resistance
MDR	-	multidrug resistance
MFS	-	major facilitator superfamily
MIC	-	minimum inhibitory concentration
MLEE	-	multi-locus enzyme electrophoresis
MLST	-	multi-locus sequence typing
NCBI	-	National Center for Biotechnology Information
NGS	-	next generation sequencing
nm	-	nanometres
NMEC	-	neonatal-meningitis/sepsis associated *Escherichia coli*
OD	-	optical density
OMP	-	outer membrane protein
PAβN	-	phenylalanine-arginine β-naphthylamide
PACE	-	proteobacterial antimicrobial compound efflux
PAP	-	periplasmic adaptor protein
PCR	-	polymerase chain reaction
RND	-	resistance-nodulation-division
SMR	-	small multidrug resistance
Spp.	-	species
ST	-	sequence type

Stx	-	Shiga toxin
TBLASTN	-	translated BLAST nucleotide
TEM	-	Temoneira
UPEC	-	uropathogenic *Escherichia coli*
UTI	-	urinary tract infection

Chapter 1 Introduction

1.1 Genome sequencing and the bacterial pangenome

The first complete bacterial genome published belonged to the human pathogen *Haemophilus influenzae.* In 1995 the process of sequencing alone required the use of 14 sequencers over a period of three months (Fleischmann *et al.*, 1995). Since then technical innovations have resulted in the reduction of both cost per base and time required to generate whole genome sequences. With the increased affordability and speed of genome sequencing, open-access genome databases have grown exponentially (Land *et al.*, 2015).

As the availability and complexity of genome sequences has increased so have the computational tools available to analyse sequence data. Using bioinformatics the presence of both antimicrobial resistance genes and secondary antimicrobial biosynthetic gene clusters can be determined without any laboratory experiments (Alcock *et al.*, 2019, Blin *et al.*, 2019, Medema *et al.*, 2011). Furthermore, phylogenetic relationships and the conservation of genes across thousands of genomes can be determined.

The term 'pangenome' was coined in 2005 and initially referred to all genes in a given study group, however it is now more commonly used to describe all genes present across a species (Tettelin *et al.*, 2005, McInerney *et al.*, 2017). Genes within a pangenome can be subdivided into core and accessory, meaning they are either present in all isolates or only a selection of isolates, respectively (Figure 1.1) (Tettelin *et al.*, 2005, Brockhurst *et al.*, 2019).

The size of a bacterial pangenome generally increases with bacterial genome size as bacteria associated with multiple niches, such as *Klebsiella pneumoniae,* have

a larger repertoire of genes across the species than organisms such as *Chlamydia trachomatis* that have evolved to be host-specific (McInerney *et al.*, 2017). One of the first pangenome analyses of *Escherichia coli* genomes found that while the mean number of genes across 17 genomes was 5,020, only ~2200 formed the core genome (Rasko *et al.*, 2008). The additional >2000 genes per genome were non-essential and varied between isolates, forming the accessory genome. A later study of 312 extraintestinal *E. coli* clinical isolates predicted a core genome size 3079 genes comprised of genes involved in basic cellular functions such as DNA replication, transcription and translation and amino acid synthesis (Salipante *et al.*, 2015).

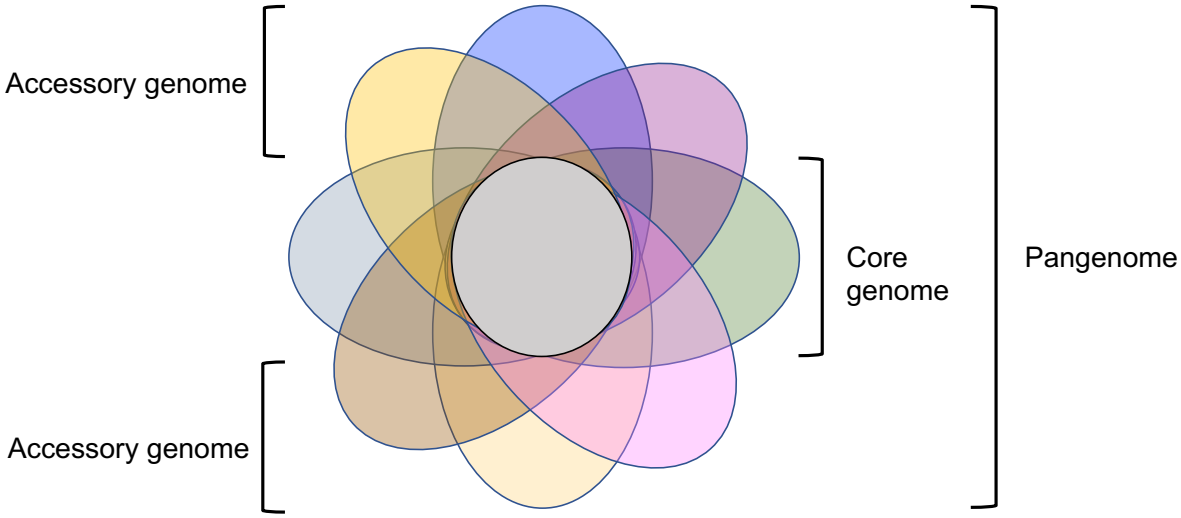


Figure 1.1 Venn diagram representation of a bacterial pangenome

The core genome here represented by the grey centre circle is made up of gene families found across all isolates within a species or population studied. Accessory genomes are formed of gene families that are present in some isolates but not all of a species or population. Figure adapted from (McInerney *et al.*, 2017).

1.2 Bacterial genome plasticity

Bacterial genomes are highly plastic allowing the fluid uptake and loss of genetic material across populations. Bacteria can acquire new fragments of DNA from the environment and neighbouring cells via horizontal gene transfer (HGT). Most HGT events are neutral or have a deleterious effect and these genes are quickly lost by the bacterium (Berg and Kurland, 2002). Beneficial genes may allow antimicrobial resistance (AMR) or survival in a novel niche and can therefore be selected for and maintained by the cell (Hall *et al.*, 2020). Modelling homologous recombination across *Campylobacter* found that increasing levels of homologous recombination increased the fixation efficiency of beneficial alleles (Woodcock *et al.*, 2017). Furthermore, non-human associated *E. coli* strains can utilise foreign DNA more easily than human-associated strains, allowing them to adapt to their environment more readily (Yang *et al.*, 2019).

Genome rearrangements can also arise via insertions, deletions, translocations and inversions. Unlike eukaryotic genomes, bacterial genomes have a deletion bias, meaning that genome reduction not genome expansion is preferred (Bobay and Ochman, 2017). Genetic drift, by which a small population size results in genome changes becoming fixed in a population, is also associated with bacterial reduction; genome reduction is also known to correlate with pathogenicity i.e. pathogenic specific isolates of a species have smaller genomes non-pathogenic (Kuo *et al.*, 2009, Conant and Wolfe, 2008, Murray *et al.*, 2021).

Insertion sequences (IS) are transposable elements present in low number across most bacterial genomes except in recently evolved pathogens such as *Shigella* species (spp.). These species contain hundreds of copies of IS elements (Touchon

and Rocha, 2007, Bobay and Ochman, 2017). Interestingly a high number of IS elements and pseudogenes across the genome is believed to occur due to the mutational hazard hypothesis whereby extreme population size reductions result in less effective gene selection (Lynch and Conery, 2003, Moran and Plague, 2004). This results in increased accumulation of non-functional sequences in early stages of genome reduction, which are eventually purged from the genome (Moran and Plague, 2004).

1.3 *Escherichia coli*

Escherichia coli is a rod shaped Gram-negative facultative anaerobe within the Enterobacteriaceae family. First identified as *Bacterium coli commune, E. coli* was described by the German microbiologist Theodor Escherich in 1884 during a study of gut bacteria from healthy infants (Escherich, 1988)[English translation]. *E. coli* is a commensal found in humans, birds and cattle but it can also be isolated from soil, sediment and water (Savageau, 1983). *Escherichia* spp. are one of the earliest colonisers of the human gastrointestinal tract and is present continuously throughout the first year of life (Wampach *et al.*, 2017). Moreover, *Escherichia* spp. have been detected in the meconium (Wampach *et al.*, 2017). Yet despite generally being a commensal organism, lineages of *E. coli* are also associated with human infection and disease.

1.4 *E. coli* infection

While most *E. coli* strains do not cause infection, those that are known to cause human disease are classified into distinct “pathovars” based on similarities in the disease

presentation due to their common virulence factors. There are six well defined intestinal pathogenic *E. coli* (InPEC) pathovars responsible for diarrhoeal disease; enteropathogenic (EPEC), enterohaemorrhagic (EHEC), enteroaggregative (EAEC), enterotoxigenic (ETEC), enteroinvasive (EIEC), and diffuse-adhering *E. coli* (DAEC). More recently an additional InPEC pathovar has been described. However adherent-invasive *E. coli* (AIEC) are not classed as diarrheagenic InPEC as they do not cause acute diarrhoea but are instead associated with inflammatory bowel conditions such as Crohn's disease (Lee *et al.*, 2019). In addition to InPEC there are also two well described extraintestinal pathogenic *E. coli* (ExPEC) pathovars; uropathogenic *E. coli* (UPEC) responsible for the majority of urinary tract infections (UTI) and neonatal meningitis/sepsis-associated *E. coli* (NMEC) which is associated with neonatal bloodborne infections (Kaper *et al.*, 2004).

UPEC strains cause infection after entering the urinary tract and attaching to the uroepithelium via pili. Cells are then internalised where they form intracellular bacterial communities, here cells replicate and produce biofilm-like structures (Anderson *et al.*, 2003). UPEC cells are able to leave these intracellular bacterial communities and re-enter the urinary tract, enabling the migration of cells up the bladder towards the kidney (Terlizzi *et al.*, 2017). Infection causes cystitis and acute pyelonephritis, resulting in tissue damage and apoptosis of bladder cells (Croxen and Finlay, 2010).

NMEC is commonly isolated from the gastrointestinal tract, however it is also the second most common cause of meningitis in newborns after group B *Streptococcus* (Gaschignard *et al.*, 2011, Weston *et al.*, 2011). Mortality rates depend on multiple factors including the age of the infant when NMEC is contracted and global

location (Nicholson *et al.*, 2016). NMEC can colonise during gestation and delivery but can also be transmitted by healthy carriers e.g. nurses (Ku *et al.*, 2015). *E. coli* enters the bloodstream and crosses the blood-brain barrier, entering the central nervous system leading to inflammation of the meninges (Croxen and Finlay, 2010).

EPEC causes infection by attaching to cells within the small intestine via bundle forming pili (Hyland *et al.*, 2008). Cytoskeletal changes such as the polymerisation of actin underneath attached cells is observed, this process is also known as attaching and effacing (A/E) (Nataro and Kaper, 1998). The ability to cause A/E is encoded on a 35 kb pathogenicity island, the locus of enterocyte effacement (LEE), the majority of which encodes a type III secretion system which translocates effector proteins from the bacterium into the host cell (Elliott *et al.*, 1998). This results in the activation of multiple kinase pathways, with downstream effects including the loosening of tight junctions between cells which results in fluid and electrolyte secretion (Kaper *et al.*, 2004).

ETEC can cause disease in both humans and animals and is endemic in low to middle income countries. ETEC can produce either heat-stable or heat-labile enterotoxins, or both, which are found primarily on plasmids (Kaper *et al.*, 2004, Croxen and Finlay, 2010). Heat-stable enterotoxins are subclassified as STa and STb with STa associated with human disease; its binding to guanylyl cyclase receptors in the small intestine results in increased levels of cyclic guanosine monophosphate and activates the cystic fibrosis transmembrane conductance regulator resulting in reduced fluid absorption (Fleckenstein and Kuhlmann, 2019, Chen *et al.*, 2015).

EHEC are primarily associated with cattle (Cray and Moon, 1995). Cells can be found throughout the digestive tract of ruminants but primarily colonise the recto-anal

junction, however infection in cattle is asymptomatic due to the lack of epithelial globotriaosylceramide (Gb3) receptors (Pruimboom-Brees *et al.*, 2000, Naylor *et al.*, 2003). Human infection requires a small infective dose of between 10-100 colony forming units (CFU) to cause infection with characteristics of disease including diarrhoea that may contain blood and haemolytic-uremic syndrome (HUS) (Kaper *et al.*, 2004, Nguyen and Sperandio, 2012). One of the key virulence factors of EHEC strains is the Shiga toxin (Stx), also called verocytotoxin. The B subunit of this toxin binds to Gb3 on the vascular epithelium, following cell entry the A subunit interacts with the ribosome, inhibiting protein synthesis (Tarr *et al.*, 2005, Melton-Celsa, 2014).

EAEC causes persistent diarrhoea in humans in both the developed and developing world and is the most common cause of travellers' diarrhoea (Kaper *et al.*, 2004). EAEC adhere to epithelial cells in the small intestine and/or colon, and aggregate into a 'stacked brick' formation; the ability of EAEC strains to produce a mucus biofilm aids persistence (Hebbelstrup Jensen *et al.*, 2014, Servin, 2014). The 2011 German *E. coli* outbreak responsible for 53 deaths was caused by EAEC/EHEC hybrid strain that had phenotypes of both pathotypes (Bielaszewska *et al.*, 2011, Mellmann *et al.*, 2011).

EIEC is highly similar to *Shigella* spp., so much so that genetically they could be one pathovar/genus; separation of the EIEC and *Shigella* is achieved through biochemical, molecular and physiological tests (Pettengill *et al.*, 2015, van den Beld and Reubsaet, 2012). Cells enter the epithelium in the colon and induce lysis of the endocytic vacuole and subsequently replicate intracellularly (Kaper *et al.*, 2004). Strains are generally non-motile, with dissemination occurring through cell-to-cell

migration, a process which is achieved through the manipulation of host cell machinery (Pasqua *et al.*, 2017).

DAEC is a heterogeneous group that generate a diffuse pattern of adherence on both HeLa and HEp-2 cells (Nataro *et al.*, 1985, Nataro *et al.*, 1987, Scaletsky *et al.*, 1984). It is further characterised by the ability to induce the formation of cellular extensions from host cells which surround bacterial cells (Kaper *et al.*, 2004). Strains of DAEC expressing Afa/Dr adhesins have been implemented in UTI infections, pregnancy complications and diarrhoea in children (Servin, 2014).

1.5 *E. coli* typing and classification

E. coli interacts with a wide range of environments and therefore needs diverse genetic capabilities, this is demonstrated by the genetic variation across populations. Prior to the development of molecular techniques serotyping was the main way to distinguish between *E. coli* isolates whereby isolates were categorised according to the agglutination reactions of surface (O-), flagellar (H-), and capsular (K-) antigens (Kauffmann, 1947). Over time, the K-antigen became less frequently used and serotyping based upon only the O- and H-antigens became the gold standard (Fratamico *et al.*, 2016). Despite the advances in molecular typing, serological typing is still widely used and some strains such as *E. coli* O157:H7 are still commonly referred to by their serological profile.

Initial *E. coli* population typing based on single genes resulted in different phylogenetic relationships depending on the genes selected (Tenaillon *et al.*, 2010). Subsequent electromorph mobility profiles using 11 enzyme loci across >1600 *E. coli* isolates resulted the assignment of three *E. coli* clusters (Ochman *et al.*, 1983). This

finding lead to the authors creating the *E. coli* reference (ECOR) collection (Ochman and Selander, 1984). The ECOR collection is a series of 72 *E. coli* strains selected in 1984 to represent the genotypic diversity across the *E. coli* population. Strain selection for the ECOR collection was based upon three criteria; strains had to have 1) been included in previous studies published by Ochman and Selander, 2) represent the three subgroups of *E. coli* which were previously determined by enzyme loci alleles and 3) represent the range of host species and geographical distribution (Ochman and Selander, 1984).

Expansion from 11 enzyme loci to 35 resulted in these three groups being further divided into six phylogenetic groups (phylogroups) A, B1, B2, C, D and E (Selander *et al.*, 1987) which were later confirmed by multi-locus enzyme electrophoresis (MLEE) a method based on enzyme polymorphisms including those found in esterases (Goullet and Picard, 1989). Since the implementation of molecular typing techniques the ECOR collection has been shown to includes strains from phylogroups A, B1, B2, C, D, E and F (Clermont *et al.*, 2015).

Developed in the late 1990's multiple locus sequence typing (MLST) uses the sequences of 'housekeeping' genes, genes responsible for the maintenance of the cell and therefore present in all isolates of the species, to separate populations (Martin *et al.* 1998). This method assigns a sequence type (ST) number to genomes depending on the combination of housekeeping gene alleles present. Currently there are several published MLST schemes in use, the most common are the Achtman Warwick and Pasteur schemes (Clermont *et al.*, 2015). Numbers assigned to sequences by these schemes differ for example, the B2 reference strain for the determination of minimal inhibition concentration (MIC) values for strains *E. coli* ATCC 25922 is ST73 under the

Achtman Warwick scheme but ST52 under the Pasteur scheme and *E. coli* O157:H7 Sakai is ST11 under the Warwick scheme but ST296 on the Pasteur scheme. Nevertheless, phylogenetic relationships between strains are highly correlated regardless of MLST system used (Gordon *et al.*, 2008). Furthermore, the phylogenetic relationships identified in MLEE are seen in phylogenetic analyses using MSLT alleles.

If gene sequencing is unavailable PCR techniques have been developed that enable the identification of phylogroups. Amplification of *chuA, yjaA, arpA* and a fragment known as TspE4.C2 allows the differentiation of A, B1, B2, C, D, E and F (Clermont *et al.*, 2000, Gordon *et al.*, 2008, Clermont *et al.*, 2013).

Phylogenetic analysis shows a divide between phylogroups A, B1, C and E, and B2, D, F (Figure 1.2), with the recently identified phylogroup G clustered with B2 strains (Clermont *et al.*, 2019). It is believed that phylogroups A and B1 have emerged more recently than B2, D, E, F and G, with phylogroup D being related closest to ancestral *E. coli* (Figure 1.3) (Touchon *et al.*, 2009, Gonzalez-Alba *et al.*, 2019). A more recent study using a Mash-based approach suggested that *E. coli* phylogroups could be further divided resulting in 14 separate phylogroups (Abram *et al.*, 2021). Mash is a programme that approximates genome similarity based upon their nucleotide content (Ondov *et al.*, 2016). Furthermore, differences in genome sizes are seen between the phylogroups. In a study of over 10,000 *E. coli* assemblies, Horesh *et al.* has demonstrated that B1, E and F isolates have the largest genomes while C, B2 and A had the smallest (Horesh *et al.*, 2021).

Though pathovars are highly useful for defining clinical presentation, molecular typing has demonstrated that pathovars can span multiple phylogroups. For example, though NMEC isolates can belong to phylogroups A or B1, the majority belong to B2

or D (Wijetunge *et al.*, 2015). However most clinical isolates responsible for causing ExPEC infections in humans belong to phylogroups B2 or D. These are phylogroups in which isolates have the greatest number of virulence factors (Clermont *et al.*, 2013, Boyd and Hartl, 1998). In contrast, commensal strains are more likely to belong to phylogroups A or B1 (Picard *et al.*, 1999).

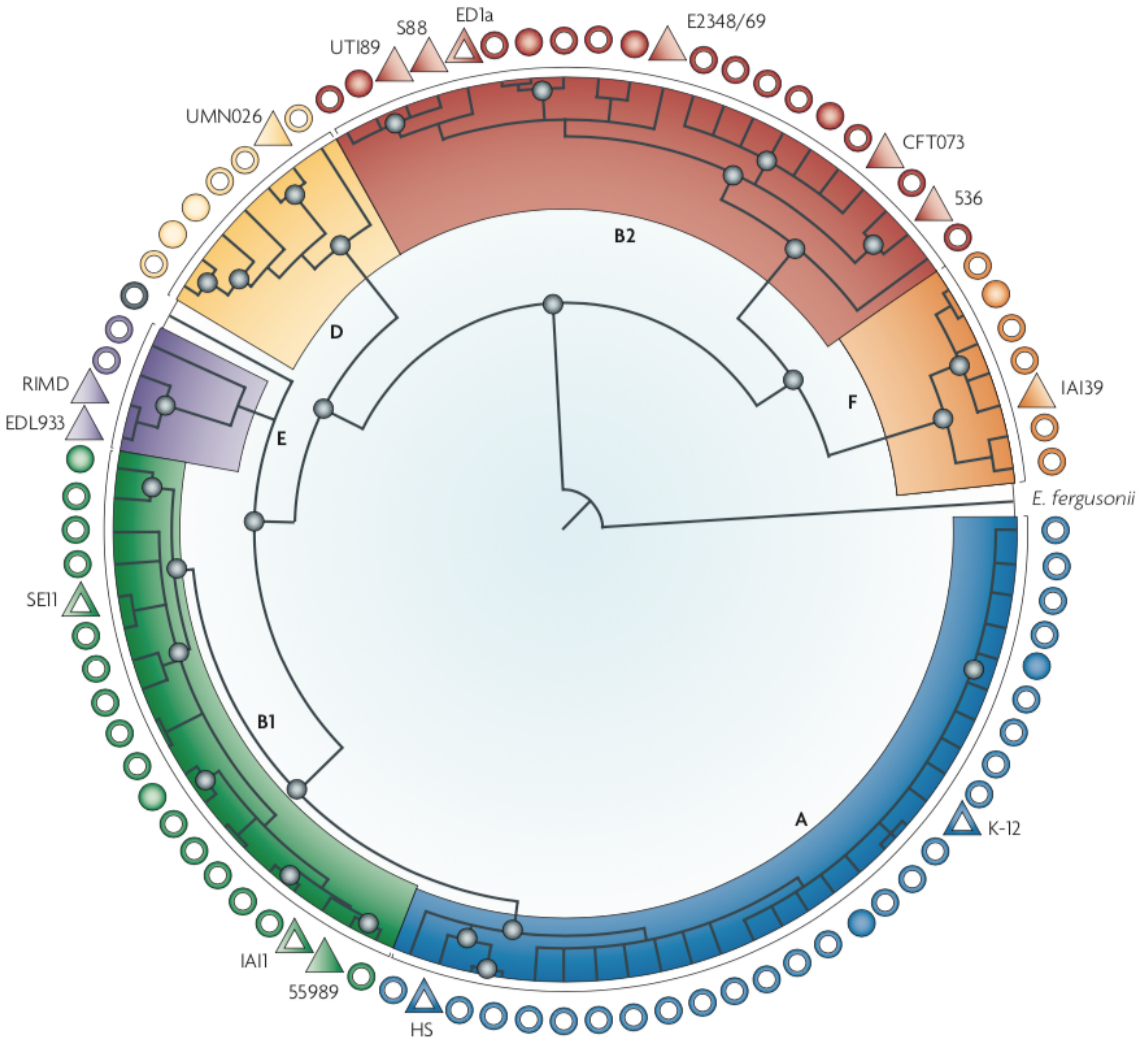


Figure 1.2 Phylogenetic tree of *Escherichia coli*

E. coli is formed of six main phylogroups (A, B1, B2, D, E, F), these are sequences grouped together due to their phylogenetic relationship. Phylogroups A and B1 associated with commensalism while B2, D and F are more associated with extraintestinal infection. Phylogroup E strains can be intestinal pathogenic *E. coli* and avian pathogenic *E. coli.* The tree shown was generated using the 72 ECOR collection stains (represented by circles) and 15 reference genomes (shown with triangles). Where the outer symbol is filled the stain is pathogenic. Figure reproduced from (Tenaillon *et al.*, 2010).

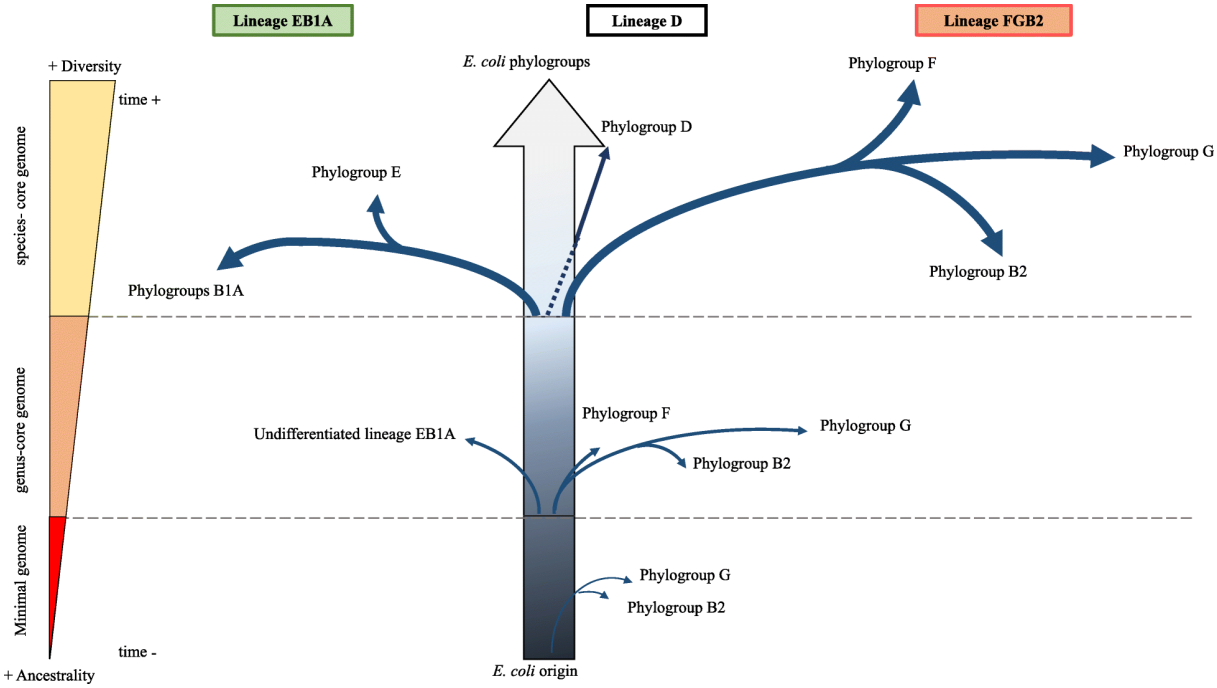


Figure 1.3 Proposed diversification of *Escherichia coli*

Proposed evolution path of *E. coli* phylogroups based upon the minimal genome, genus-specific genome, and species-specific genome. Genes involved in carbohydrate utilisation are more prevalent in phylogroups A, B1 and E, while genes involved in environment adaption were absent in phylogroups B2, F and G. It is suggested that phylogroup D isolates have genomes most like those of ancestral *E. coli* as it contains the fewest single nucleotide polymorphisms when compared to the most recent common ancestor. Figure reproduced from (Gonzalez-Alba *et al.*, 2019).

1.6 Mechanisms of resistance

Antimicrobial resistance (AMR) genes have existed since microorganisms evolved the ability to produce secondary metabolites (Benveniste and Davies, 1973, Perry *et al.*, 2016). However, today AMR is a global concern clinically. In 2014 a UK government report predicted that by 2050 up to 10 million people a year could die due to AMR, although the exact number has been contested (O'Neill, 2014, de Kraker *et al.*, 2016).

AMR can either be intrinsic or acquired, meaning it can be due the innate cellular physiology or due to the acquisition of mobile genetic elements, respectively. Moreover, AMR occurs through four main mechanisms (Figure 1.4); 1) modification of the drug target 2) enzymatic degradation or modification of the drug, 3) metabolic bypass, 4) reduced uptake or increased efflux (Blair *et al.*, 2015c).

Target modification is one of the many mechanisms by which bacteria evade antimicrobials. Here the direct target of the antimicrobial is altered, preventing the interaction with the drug. Target modification in *Staphylococcus aureus* reduces susceptibility to vancomycin and can occur through increasing the thickness of the peptidoglycan layer, making it more difficult for the drug to interfere with cell wall biosynthesis, or the modification of the amino acid sequence in the peptidoglycan peptide chain (Cui *et al.*, 2003). For example, after changing D-ala-D-ala to D-ala-D-lac or D-ala-D-ser, vancomycin is no longer able to prevent the formation of the "D-ala-D-ala" link, thus crosslinking within peptidoglycan can occur (Courvalin, 2006).

In 2015 enzymes that modify the outer membrane and therefore interrupt the mode of action for colistin were found mobilised on a plasmid for the first time (Liu *et al.*, 2016). Prior to this resistance to the last resort polymyxin was associated with chromosomal mutations. This mobile colistin resistance (*mcr-1*) gene was first

identified from *E. coli* isolated from a pig farm (Liu *et al.*, 2016) and since 2015 a further eight *mcr* variants have been detected globally (Ling *et al.*, 2020). The *mcr* genes encode phopshoethanolamine transferases, these enzymes result in a reduction of negative charge for lipid A, subsequently reducing the negative charge of the outer membrane preventing cell membrane disruption by colistin (Hinchliffe *et al.*, 2017, Ling *et al.*, 2020, Li *et al.*, 2019a).

The expression of a range of enzymes results in the degradation or modification of antimicrobials, reducing their activity and/or their availability. Enzymatic modification of the drug can occur through multiple reaction types including hydrolysis and group transfer e.g. acyl, thiol, glycosyl (Wright, 2005). A classic example of enzymatic drug modification is the inactivation of β-lactams and cephalosporins by β-lactamase enzymes. These enzymes, frequently mobilised on plasmids, work by hydrolysing the amide bonds of the β-lactam ring and were first reported in 1940 (Tooke *et al.*, 2019, Abraham and Chain, 1940).

There are numerous types of β-lactamases including narrow-range β-lactamases (also called penicillinases), extended spectrum β-lactamases (ESBL), metallo-β-lactamases, carbapenemases (also called *Klebsiella pneumoniae* carbapenemases - KPC), and cephalosporinases. The majority of ESBLs can be divided into three main groups, TEM (Temoneira), CTX-M (a cefotaximase isolated first in Munich) and SHV (sulfhydryl variable) (Pitout and Laupland, 2008). A recent review of the literature suggested that carriage of ESBL-*E. coli* within the community averaged at 16.5%, however continental carriage rates ranged considerably (Bezabih *et al.*, 2021). In Europe carriage was at 6.0%, Eastern Mediterranean 20.6%, Africa 21.4%, Western Pacific 24.5% and carriage in South-East Asia was at 27.0% (Bezabih

et al., 2021). Treatment of ESBL infections must be designed to minimise unnecessary use of carbapenems due the risk of selecting for resistance (Bitsori and Galanakis, 2019). KPC-producing Enterobacteriaceae isolates are only susceptible to colistin and tigecycline; some isolates are also susceptible to amikacin and gentamicin (Nordmann *et al.*, 2009). The same limited susceptibility is seen for metallo-β-lactamase producing isolates as most are only sensitive to colistin and tigecycline (Kumarasamy *et al.*, 2010). However, tigecycline and colistin resistance is an increasing threat.

Tigecycline, like colistin, is a last resort drug. Resistance has been shown to occur through both enzymatic inactivation and efflux. The *tetA*, *tetM* and *tetX* genes encode a tetracycline efflux pump, a ribosomal protection protein, and monooxygenase that results in the breakdown of all known tetracyclines (including tigecycline), respectively (Yang *et al.*, 2004). TetX was first identified on the chromosome of the anaerobe *Bacteroides fragilis* but has since been detected in clinical isolates of *Acinetobacter baumannii* and Enterobacteriaceae (Leski *et al.*, 2013). Moreover, the mobilised *tetX* allele *tet(X4)* has been detected in *E. coli* samples in China (Sun *et al.*, 2019a).

Metabolic bypass is when the original antimicrobial target is replaced. An example of which is the acquisition of the penicillin binding protein 2a (PBP2a) by methicillin resistant *Staphylococcus aureus.* PBP2a is a non-*S. aureus* transpeptidase involved in the crosslinking of peptidoglycan. Encoded by *mecA,* PBP2a has a reduced affinity to β-lactams than the native PBP thus allowing cell wall biosynthesis to continue despite exposure to β-lactams (Fishovitz *et al.*, 2014).

The outer membrane of Gram-negative bacteria provides intrinsic protection against antimicrobials due to reduced cell permeability and they are therefore less

susceptible to antimicrobials than Gram-positive bacteria. Hydrophilic antibiotics can enter the periplasm through porins located in the outer membrane and pass across the inner membrane (Blair *et al.*, 2015c). Through the downregulation of porin genes, *E. coli* and other Gram-negative bacteria can further reduce their cell permeability and the uptake of such molecules (Delcour, 2009). Porins have also been shown to modulate membrane integrity, reducing permeability (Choi and Lee, 2019). Additionally Gram-negative bacteria have the ability to upregulate expression of efflux pumps that span the inner membrane. These proteins actively transport molecules from the cytoplasm into the periplasm or from the periplasm into an associated outer membrane protein (OMP). As many efflux pumps have the ability to form complexes with additional proteins, molecules can be exported directly from the periplasm into the surrounding environment. Upregulation of the efflux pumps can occur through mutations in their regulators or transiently as a response to environmental stimuli.

Moreover, cells can possess a combination of resistance genes, providing them with an armoury against even the last resort drugs. In *E. coli* resistance to both tigecycline and colistin has been reported due to the presence of *mcr-1* and mutations in the efflux pump AcrB as well as the post-transcriptional regulator Lon (Wang *et al.*, 2018).

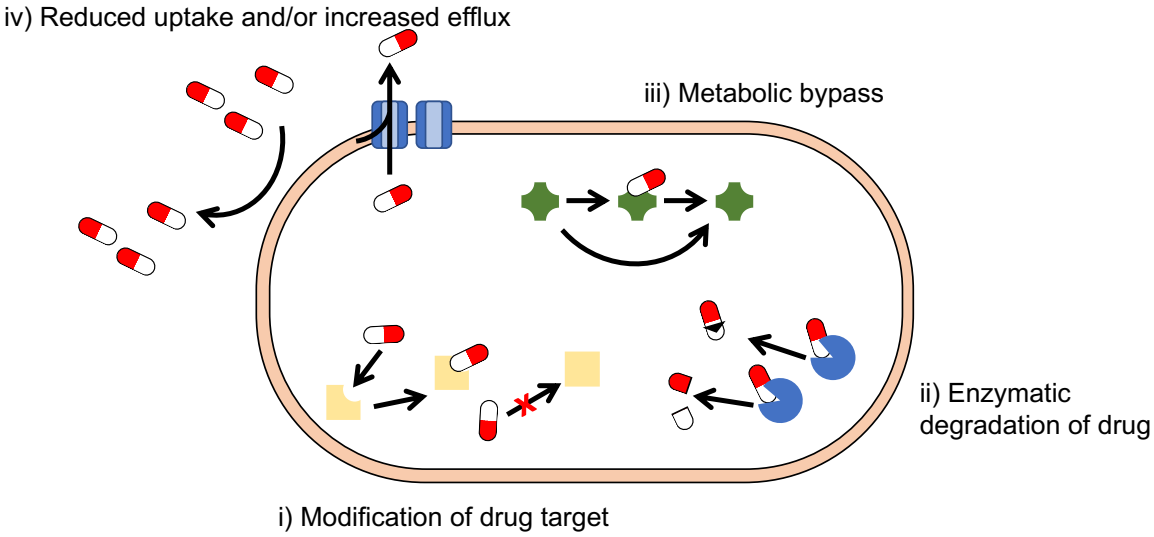


Figure 1.4 Mechanisms of resistance

Escherichia coli can evade the bacteriostatic and bactericidal effects of antimicrobials via multiple mechanisms i) modification of the drug target preventing drug interaction ii) enzymatic modification and degradation of the drug iii) metabolic bypass and iv) reduced uptake and increased efflux though porin and efflux pump expression. Figure was adapted from (Zohra *et al.*, 2021).

1.7 Multidrug resistance through efflux

Efflux pumps are present in all domains of life and are classified by their protein sequence. In bacteria, there are six major classes of pump that can efflux antimicrobials, the major facilitator superfamily (MFS), the small multidrug resistance family (SMR), the ATP-binding cassette superfamily (ABC), the multidrug and toxic compound extrusion family (MATE), the proteobacterial antimicrobial compound efflux family (PACE) and the resistance-nodulation-division family (RND) (Figure 1.5) (Hassan *et al.*, 2019, Li *et al.*, 2015). All use the proton motive force to pump compounds from the periplasm and inner membrane across the inner membrane, except for the ABC family which uses adenosine triphosphate (ATP).

MFS transporters are located in the inner membrane and transport molecules from the cytosol to the periplasm (Li *et al.*, 2015). While MFS pumps can function as standalone proteins to transport substrates across the inner membrane, many form tripartite complexes, such as EmrAB-TolC in *E. coli,* allowing transport of substrates directly from the cytoplasm out of the cell (Tanabe *et al.*, 2009). Overexpression of MFS transporters has been associated with a multidrug resistant (MDR) phenotype. For example, MdfA in *E. coli*, has been reported to increase tolerance to a broad range of substrates including chloramphenicol, tetracycline, rifampicin and erythromycin (Edgar and Bibi, 1997). In *A. baumannii,* the MFS pump CraA has been demonstrated to decrease susceptibility to benzalkonium, norfloxacin, chloramphenicol, and chloramphenicol derivatives (Roca *et al.*, 2009, Foong *et al.*, 2019).

ABC pumps are located in the inner membrane and are the only group of efflux pumps that use ATP hydrolysis as the energy source for molecule transport across the inner membrane. One of the best characterised ABC exporter pumps is MacB in *E.*

coli, which has also been shown to form a multiprotein complex with MacA and TolC. The expression of MacAB-TolC has been shown to decrease macrolide, colistin and bacitracin susceptibility (Crow *et al.*, 2017, Kobayashi *et al.*, 2001). Furthermore, in *E. coli* MacAB exports enterotoxin II, a heat-stable toxin produced by ETEC (Yamanaka *et al.*, 2008). Increased expression of *macAB* has been reported in clinical isolates of *A. baumannii* following tigecycline exposure (Lin *et al.*, 2017).

SMR pumps are associated with resistance to quaternary ammonium compounds (QAC) that are used in disinfectants, such as benzalkonium chloride (Merchel Piovesan Pereira and Tagkopoulos, 2019, Bay *et al.*, 2008). Some *Enterococcus faecalis* isolates possess QacZ, an SMR pump that is involved in QAC tolerance (Braga *et al.*, 2011). SMR pumps have also been linked to antimicrobial susceptibility in *A. baumannii,* deletion of SMR transporter *abeS* reduced the minimum inhibitory concentrations of chloramphenicol, ciprofloxacin and erythromycin (Srinivasan *et al.*, 2009).

MATE pumps are present in both eukaryotes and prokaryotes and may utilise both the proton and sodium motive forces to export molecules across the inner membrane (Ogawa *et al.*, 2015). All MATE pumps appear to have the ability to efflux fluoroquinolones and many also export the DNA-chelating dye ethidium bromide (Kuroda and Tsuchiya, 2009). The MATE pump EmmdR in *Enterobacter cloacae* has been linked to increased fluoroquinolone resistance (He *et al.*, 2011) and in a laboratory strain of *Staphylococcus aureus,* MepA has been shown to be involved in reduced tigecycline susceptibility (McAleese *et al.*, 2005).

PACE pumps were the most recently identified family of efflux pumps and the best characterised example is AceI from *A. baumannii*. Located in the inner membrane,

AceI has been shown to efflux the biocide chlorhexidine and homologs have been found in many other bacterial species though predominantly those that belong to the phylum proteobacteria, including *E. coli* (Hassan *et al.*, 2013, Hassan *et al.*, 2018, Hassan *et al.*, 2019). However, no PACE system in *E. coli* has yet been characterised, nor has the structure of PACE pumps been resolved.

RND pumps are exclusively found in Gram-negative bacteria, and form tripartite complexes spanning the cytoplasmic membrane, periplasmic space and outer membrane (Figure 1.6). As a result, molecules extruded by RND pumps are less likely to re-enter the cytosol through free diffusion than those removed by other classes (Li *et al.*, 2015). RND pumps confer resistance to a wide range of compounds and while removal of any component of an RND pump renders a system non-functional, promiscuity has been demonstrated for PAPs in *S.* Typhimurium, meaning PAP proteins can form complexes and function with multiple RND pumps (Smith and Blair, 2014, Lau and Zgurskaya, 2005).

Substrate specificity of efflux pumps can vary greatly. Pumps can have a narrow substrate range like the MFS family pump TetA, which extrudes tetracycline, or very broad ranges like AcrB which can transport structurally diverse substrates (Li *et al.*, 2015). Overexpression of efflux pumps, specifically RND pumps, is associated with MDR in both the laboratory and clinical settings. Clinical isolates that overexpress *acrB* have reduced fluoroquinolone and carbapenem susceptibility (Swick *et al.*, 2011, Chetri *et al.*, 2019) and increased expression of the RND operon *mdtEF* was found to correlate with AMR in clinical isolates of *E. coli* (Yasufuku *et al.*, 2011). In *A. baumannii* overexpression of AdeABC through mutations in the two-component system AdeRS has been described in clinical MDR isolates, resulting in increased resistance to a

range of substances including aminoglycosides, fluoroquinolones, tetracycline, tigecycline (Yoon *et al.*, 2013). Notably, it has been shown in *S.* Typhimurium that drug susceptibility can be restored through both the deletion of *acrB* and the loss of *acrB* function (Wang-Kan *et al.*, 2017).

Efflux pumps are generally encoded chromosomally, however their genes can be mobilised on transmissible elements like plasmids (e.g. *qnr* genes on pMG252 (Martinez-Martinez *et al.*, 1998) and transposons (e.g. Tet pumps on Tn*10*). AMR through efflux is most commonly mediated through chromosomal mutations in transcriptional regulators resulting in increased expression of the pump. However, the chromosomally encoded *Klebsiella pneumoniae* pump OqxAB, which exports fluoroquinolones has been found on the plasmid pOLA52 in *E. coli,* with *oqxAB* flanked by IS*26*-like sequences (Norman *et al.*, 2008). Originally isolated from swine manure and conferring resistance to chloramphenicol and ampicillin, in addition to olaquindox, the OqxAB pump has been identified in human samples (Hansen *et al.*, 2004, Hong *et al.*, 2009, Sørensen *et al.*, 2003).

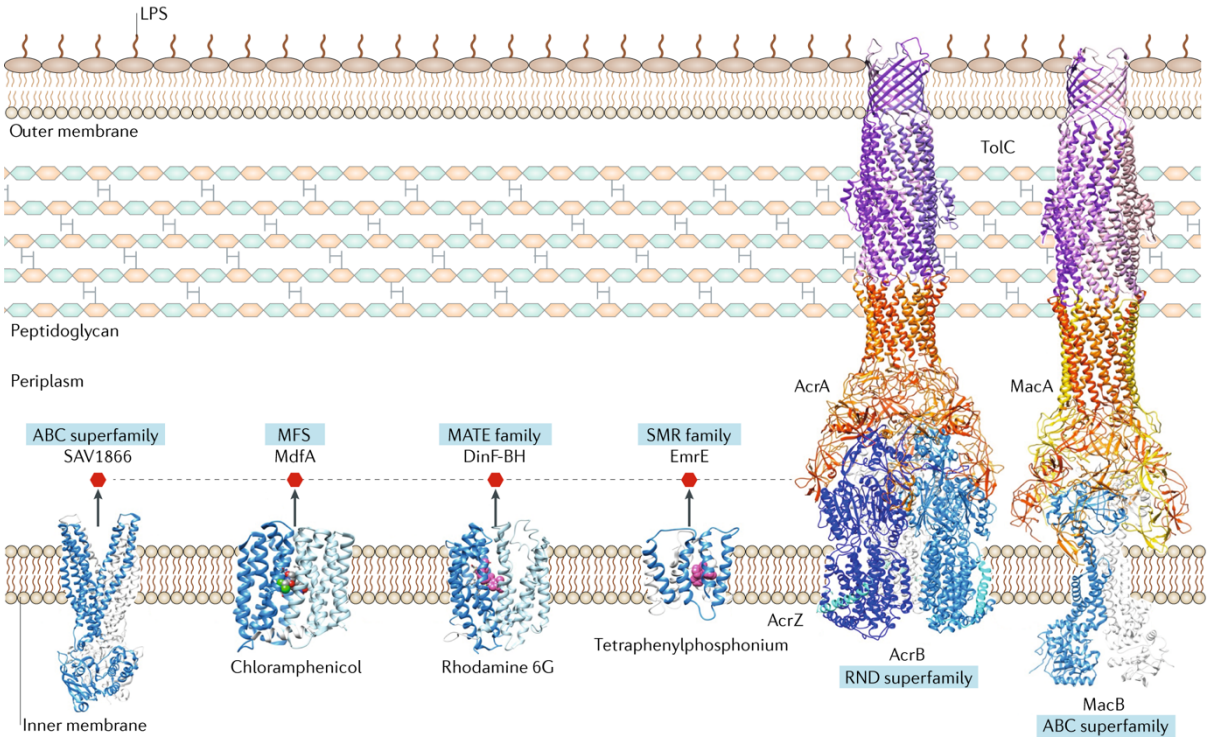


Figure 1.5 Classes of drug efflux pumps present in *E. coli*

There are six families of efflux pumps associated with the transport of antimicrobials present in *E. coli*. The ATP-binding cassette superfamily (ABC), major facilitator superfamily (MFS), multidrug and toxic compound extrusion family (MATE), small multidrug resistance family (SMR), proteobacterial antimicrobial compound extrusion (PACE) (not shown) and the resistance-nodulation-division (RND) division. Figure from (Du *et al.*, 2018).

1.8 RND systems in *E. coli*

It is widely accepted that *E. coli* isolates have six RND efflux pumps, AcrB, AcrD, AcrF, CusB, MdtBC, and MdtF, however much of the work done to characterise efflux pumps and their functions has been done in the laboratory strain *E. coli* K-12 substrain MG1655 which is known to have a genome considerably smaller than *E. coli* strains isolated from infections (Hayashi *et al.*, 2001).

The best characterised RND pump in *E. coli* and across the Enterobacteriaceae is AcrB (Figure 1.6). In *E. coli,* this pump is expressed under laboratory conditions and has a broad substrate range that includes clinical antimicrobials such as chloramphenicol, tetracyclines and fluoroquinolones, biocides including triclosan, dyes such as ethidium bromide, and also the iron-chelator enterobactin, which binds extracellular iron and transports it across the membrane into the cellular periplasm (Horiyama and Nishino, 2014, Buchanan *et al.*, 1999, Rutz *et al.*, 1992, Nishino and Yamaguchi, 2001, McMurry *et al.*, 1998). In *S.* Typhimurium deletion of *acrB* has been shown to attenuate virulence in both mice and cellular models of infection, and reduce biofilm production (Nishino *et al.*, 2006, Baugh *et al.*, 2014).

There is a level of redundancy seen between efflux pumps i.e. the function of a protein compensates for the loss of another. This is particularly true for AcrF which is a close homolog to AcrB and has an overlapping, though narrower, substrate range with AcrB. These substrates include the dye ethidium bromide, fluoroquinolones, β-lactams, tetracyclines and indole (Kawamura-Sato *et al.*, 1999, Nishino and Yamaguchi, 2001, Nishino *et al.*, 2003, Alav *et al.*, 2021). In *E. coli* the *acrEF* operon is repressed by H-NS (Nishino and Yamaguchi, 2004).

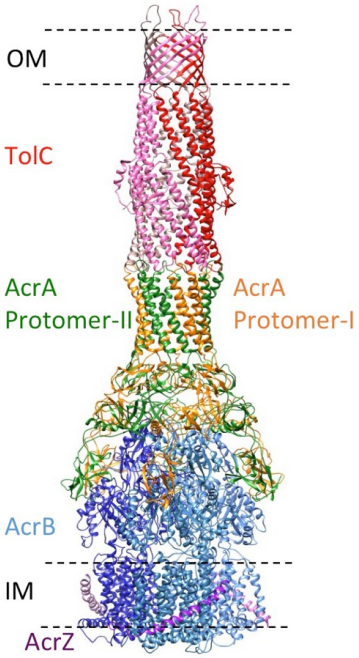


Figure 1.6 The structure of AcrAB-TolC.

Drugs enter the AcrAB-TolC complex in the periplasm or through the outer leaflet of the inner membrane through AcrB and are extruded through the AcrA hexamer. AcrA is a periplasmic adapter, linking AcrB to the outer membrane channel, TolC (Wang *et al.*, 2017).

Like *acrB,* interruption of the *acrF* gene results in impacted virulence in the BALB/c mice model (Nishino *et al.*, 2006). *In vitro* models have also shown reduced adhesion and invasion when *acrF* in *Salmonella* is disrupted (Blair *et al.*, 2015b).

AcrD has a narrower and more distinct substrate range than AcrB and AcrF, as it has been shown to export aminoglycosides and anionic β-lactams (Nishino *et al.*, 2003). In the presence of chloramphenicol *E. coli* lacking *acrB* have been shown to increase *acrD* transcription and induction has also been seen in the presence of indole (Cudkowicz and Schuldiner, 2019, Hirakawa *et al.*, 2004). AcrD, like AcrB, forms a tripartite complex with the AcrA PAP and TolC OMP (Yamasaki *et al.*, 2011). Furthermore, like AcrB, AcrD has been shown to export enterobactin (Horiyama and Nishino, 2014). In *S.* Typhimurium, AcrD has been linked to virulence with inactivation of the *acrD* gene resulting in decreased pathogenicity in mice (Nishino *et al.*, 2006). However strains with inactivated *acrD* have also been shown to outcompete the parental *S.* Typhimurium SL1344 strain (Buckner *et al.*, 2016).

MdtBC is the only RND pump in *E. coli* that is formed of a heterotrimer and has a B_2C stoichiometry (Kim *et al.*, 2010). This two-protein unit is believed to have arisen through a gene duplication event (Gorecki and McEvoy, 2020). It is not fully understood how the two subunits come together as it has been previously suggested that only MdtB can function independently by one group, yet another group has shown that only MdtC can function independently (Kim *et al.*, 2010, Wang and Fierke, 2013). MdtABC has been shown to export enterobactin in *E. coli;* sodium tungstate and flavonoids are also believed to be substrates (Horiyama and Nishino, 2014, Leblanc *et al.*, 2011). It has been shown to be regulated by the two-component system BaeRS which is involved in cellular responses to environmental stress (Baranova and Nikaido, 2002,

Raffa and Raivio, 2002). Furthermore both BaeRS and MdtABC have been linked to zinc tolerance (Wang and Fierke, 2013). A homologous system is present in *Erwinia amylovora* where overexpression of MdtABC reduced susceptibility to bile salts and silver nitrate (Pletzer and Weingart, 2014).

Like MdtABC, researchers have only recently begun to elucidate the role of MdtEF in *E. coli.* MdtEF is flanked by *gadEF* downstream and *gadWYX* upstream and has been shown to be regulated by GadX, an AraC-family regulator involved in acid tolerance (Nishino *et al.*, 2008a, Giangrossi *et al.*, 2005). The cyclic AMP receptor protein (CRP) has also been shown to regulate *mdtEF* expression (Nishino *et al.*, 2008b) and increased expression is seen in both anaerobic and nitrosative environments (Zhang *et al.*, 2011, Deng *et al.*, 2013). Furthermore, MdtE provides a fitness benefit during periods of extreme acid exposure (Schaffner *et al.*, 2021). These findings imply MdtEF-TolC may provide a level of protection for *E. coli* when in hostile environments such as the human host.

CusABC has the narrowest substrate range of the RND systems, exporting copper and silver ions (Franke *et al.*, 2001, Franke *et al.*, 2003). The periplasmic element of this pump, CusB, has been shown to be key in sensing metal ions (Santiago *et al.*, 2017). The efflux pump complex is associated with the non-essential *cusF,* this encodes a periplasmic metallochaperone that is believed to deliver copper and silver ions to the CusABC system (Xue *et al.*, 2008, Kittleson *et al.*, 2006). Deletion of *cusF* was found to increase silver susceptibility in a silver resistant *E. coli* strain (Lok *et al.*, 2008). The *cus* system is an anaerobic copper tolerance system (Outten *et al.*, 2001). A 2014 study found that UPEC grown in urine from UTI patients had increased expression of *cusC* and the *cus* operon regulator *cusR* when compared to samples

grown in urine from healthy women and lysogeny broth, moreover interruption of the *cus* system (*cusSRCFBA::aph*) was found to result in enhanced susceptibility to copper sulphate in anaerobic conditions (Subashchandrabose *et al.*, 2014). It is suggested that this system has a significant role in UPEC colonisation and survival as a result.

1.9 AcrAB-TolC

AcrB is an RND efflux pump that forms a complex with AcrA and TolC resulting in an efflux system that spans the inner membrane, periplasm and outer membrane with a 3:6:3 stoichiometry (AcrA:AcrB:TolC) (Figure 1.6) (Du *et al.*, 2014). The AcrB protein is a homotrimer, switching between an inactive symmetrical and active asymmetrical configuration (Yamaguchi *et al.*, 2015). While in the asymmetrical conformation, AcrB is hypothesised to export substances through a three-step mechanism called functional rotation (Figure 1.7) (Murakami *et al.*, 2006). During this process, each protomer cycles through three conformations; access, binding and extrusion, these are also referred to as L, T and O respectively (Figure 1.7) (Wang *et al.*, 2017). When in the L conformation, the vestibule is open to the periplasm providing access to potential substrates (Murakami *et al.*, 2006). After moving to T the vestibule remains open but the binding pocket also opens up and adjusts in response to entry of the substrate (Murakami *et al.*, 2006). In the third conformation, O, the vestibule is closed and exit opened, the bound molecule is then pushed out due to the shrinking of the binding pocket (Murakami *et al.*, 2006, Wang *et al.*, 2017) and removed from the periplasmic space via the outer membrane channel, TolC. Recent studies have shown that there

are four channels for drug entry into AcrB with differing substrate specificities (Tam *et al.*, 2021).

AcrZ is a small peptide which associates with AcrB at the inner membrane and has been shown to modify AcrB activity in *E. coli* (Hobbs *et al.*, 2012). Deletion of *acrZ* increased sensitivity of *E. coli* to tigecycline, chloramphenicol and puromycin (Hobbs *et al.*, 2012). Furthermore, *acrZ* transcription appears to be affected by transcriptional activators associated with an MDR phenotype, such as MarA (Hobbs *et al.*, 2012). However *acrZ* is not an essential gene in *E. coli* K-12 (Goodall *et al.*, 2018).

AcrA is a hexamer formed of alternating monomers, protomer-I and protomer-II (Wang *et al.*, 2017). Within each protomer the membrane proximal domain interacts with AcrB and a short helix-turn-helix (HTH) motif interacts in a tip-to-tip manner with HTH motifs present in TolC (Wang *et al.*, 2017). During the transport process TolC remains in an open conformation, however in isolation it assumes a closed state (Du *et al.*, 2014).

A cross sectional reconstitution of the pump complex showed an open channel from the periplasmic domain of AcrB through AcrA and TolC, to the external environment (Du *et al.*, 2014). As mentioned, in its apo-form, TolC is in a closed state, however as a result of repackaging of AcrA domains, iris-like dilation occurs at the periplasmic end of TolC (Wang *et al.*, 2017, Du *et al.*, 2018).

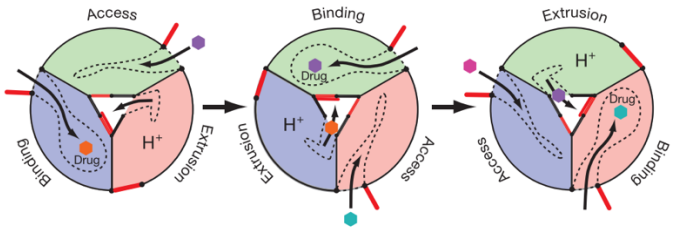


Figure 1.7. Functionally rotating mechanism of transport by AcrB

The drug enters AcrB while in the access conformation (L), next the binding pocket opens up to account for the drug (T), finally the drug is extruded from the protein in the extrusion conformation (O) (Murakami *et al.*, 2006)

1.10 AcrAB-TolC regulation

Regulation of AcrAB-TolC is complex and occurs at both the transcriptional and post-transcriptional levels (Figure 1.8). The direct transcription of *acrAB-tolC* is controlled both locally and globally by regulators. Local regulators are encoded adjacent to the pump and are often operon specific while global regulators are encoded elsewhere on the genome and control multiple promoters.

The *acrAB* operon is directly regulated by AcrR, a TetR family repressor which is divergently transcribed from *acrAB* (Ramos *et al.*, 2005). While AcrR does not completely prevent the expression of *acrAB,* it provides a level of negative regulation for both *acrR* and *acrAB* (Ma *et al.*, 1996).

The major driver of *acrAB* expression in *E. coli* is MarA which is a member of the AraC/XylS family of transcriptional activators. This family of regulators possess two HTH motifs which are fundamental for DNA-protein interaction at the major groove (Cohen *et al.*, 1993). MarA is encoded by the multiple antibiotic resistance (mar) locus which is comprised of *marA, marB, marC* and *marR*, though the locus is formed of two transcriptional units, *marC* and *marRAB* (Barbosa and Levy, 2000, Li *et al.*, 2015).

Over 10,000 marboxes are present across the *E. coli* genome and MarA has been shown to activate over 60 genes by binding to one of these regions (Sharma *et al.*, 2017). A marbox is an asymmetrical 20 bp degenerate region upstream of target operons. The transcriptional regulator *marR,* binds to *marRAB* and self-represses the operon (Ariza *et al.*, 1994). The induction of *marA* can be achieved through exposure to phenolic compounds; here MarR binds the compound, releasing from the marbox allowing transcription to proceed (Alekshun and Levy, 1999). This subsequently leads to increased expression of *marA* and consequently *acrAB,* as it is located downstream

of a marbox (Gillette *et al.*, 2000, Cohen *et al.*, 1993). Increased expression of regulators can have a direct impact on MDR due to increased expression of efflux pumps. For example, increased tigecycline resistance in *K. pneuomoniae* has been described through increased MarA expression which subsequently increased *acrB* transcription (Park *et al.*, 2020).

In addition to MarA, regulation through binding at the marbox is possible by both Rob and SoxS, both of which are also AraC/XylS transcriptional activators (Cohen *et al.*, 1993, Skarstad *et al.*, 1993). SoxS is part of the *soxRS* operon with *soxR* and *soxS* transcribed divergently, and each contain HTH motifs and can therefore bind DNA (Wu and Weiss, 1991). SoxR is a globular homodimer that responds to environmental signals enabling SoxS to activate transcription of target regulon, including *acrAB* (Wu *et al.*, 1995). The *soxRS* regulon is part of the oxidative stress response and is induced by nitric oxide, superoxide anion radicals and superoxide generators such as paraquat (Gaudu *et al.*, 1997). Rob is believed to lead to increased efflux via AcrAB and unlike MarA and SoxS, it is constitutively expressed (Rosenberg *et al.*, 2003). However, Rob molecules are usually sequestered and in a non-functional state within the cell (Griffith *et al.*, 2009). Activation of *rob* occurs in the presence of bile salts and fatty acids (Rosenberg *et al.*, 2003).

In *E. coli*, the binding of MarA, SoxS or Rob to the marbox leads to increased expression of *acrAB,* leading to an increased level of efflux and an MDR phenotype (Ricci *et al.*, 2014). While there are unique regulatory systems for each transcriptional activator, there is also some interplay between the three systems.

In *S.* Typhimurium both MarA and SoxS and are known to be regulated post-transcriptionally by Lon, an ATP-dependent protease of the AAA+ superfamily (Ricci

et al., 2014). The holoenzyme of Lon is a homo-oligomer which associates into a hexameric ring-structure (Lee and Suzuki, 2008). Each Lon domain consists of an amino-terminal domain involved in substrate binding, a proteolytic active site in the carboxyl-terminal domain, a substrate sensor domain, and an ATPase domain (Lee and Suzuki, 2008). The rapid degradation of transcriptional activators by Lon enables the cell to quickly respond to environmental changes and reduce expression levels of genes involved in a range of cellular functions, including efflux, providing a post-transcriptional level of regulation.

SdsR is one of the most abundant sRNA molecules in Enterobacteria and is present at high levels during stationary phase (Frohlich *et al.*, 2012). Overexpression of *sdsR* in *E. coli* increased susceptibility to quinolones and SdsR has been shown to repress *tolC* expression and reduce biofilm production (Parker and Gottesman, 2016, Kim *et al.*, 2015). SdsR interacts Hfq, an RNA binding chaperone protein that effects transcription indirectly (Le Derout *et al.*, 2010).

Deletion of *hfq* results in increased rhodamine 6G accumulation, increased susceptibility to antimicrobials including nalidixic acid, novobiocin, chloramphenicol and oxacillin, and reduction in AcrB protein production (Yamada *et al.*, 2010). However, no direct link between Hfq and efflux has been shown. Another sRNA, DsrA has been shown to induce the *mdtEF* operon and reduce susceptibility to a range of antimicrobials in *E. coli* (Nishino *et al.*, 2011).

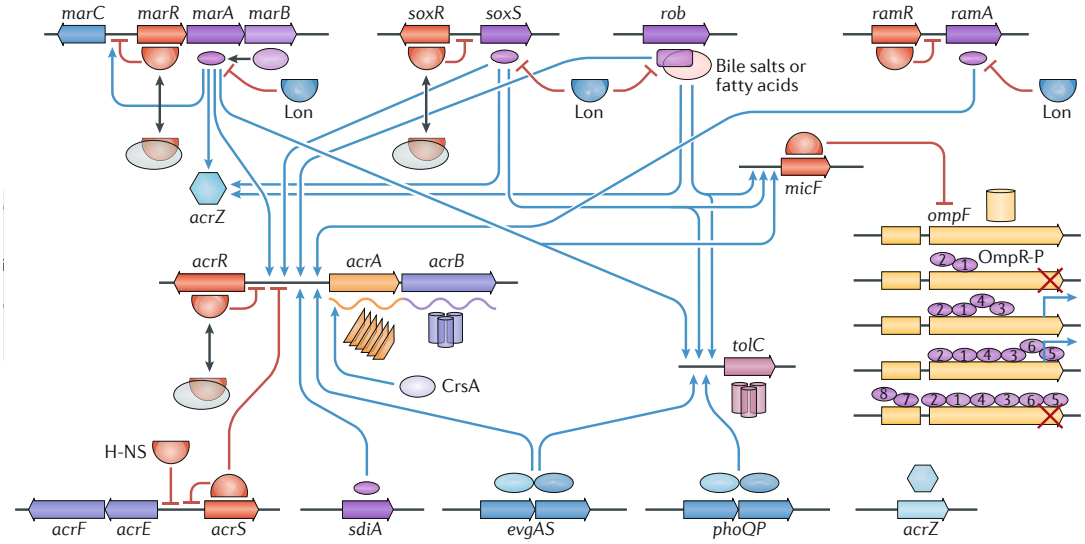


Figure 1.8 Regulation of AcrAB-TolC in Enterobacteriaceae

The regulation of the efflux pump, AcrAB-TolC is highly complex. Above shows the transcriptional factors known to act on *acrAB-tolC* in *E. coli*. Blue indicates activation, while red shows repression. *S.* Typhimurium and *K. pneumoniae* have a further AraC/XylS activator, RamA, not present in *E. coli* (Du *et al.*, 2018).

1.11 The wider role of RND efflux pumps

Efflux pumps are not only involved in AMR but are evolutionarily ancient and have fundamental physiological roles. Removing efflux associated genes or abolishing their function has been shown to have an effect on virulence and tolerance of host derived antimicrobials. For example, inactivation of *acrB* was found to increase both mouse and *Galleria mellonella* survival rates after *S.* Typhimurium infection, while invasion of intestinal epithelial cell lines *in vitro* was found to be significantly reduced (Wang-Kan *et al.*, 2017). RND efflux pumps have also been shown to export human antimicrobial peptides. The *Neisseria gonorrhoeae* RND pump, MtrCDE, has been shown to decrease susceptibility to multiple types of antimicrobial peptides including LL-37 (Shafer *et al.*, 1998).

Bile acids and bile salts are produced in the liver and have detergent properties that can cause bacterial membrane disruption. Bacteria found in the gastrointestinal tract such as *E. coli* must tolerate these molecules; one method used is efflux. In *E. coli,* AcrAB reduces bile accumulation (Thanassi *et al.*, 1997) and *acrAB* gene expression is induced in the presence of bile salts (Rosenberg *et al.*, 2003). In *A. baumannii,* AdeIJK has been shown to directly efflux human host fatty acids as well as modulate lipids in the bacterial membrane (Jiang *et al.*, 2019).

Deletion of efflux pump components has also been shown to affect biofilm production. Baugh *et al.* (2014) showed that inactivating *acrB* and *tolC* reduced biofilm production in *S.* Typhimurium ATCC 14028s through reduced curli production (Baugh *et al.*, 2014). Curli is a proteinaceous fibre produced by various bacterial species including *E. coli* and *Salmonella spp.* that is involved biofilm formation (Barnhart and Chapman, 2006).

Outside of the clinical setting, efflux pumps export antimicrobials produced by plants and soil-dwelling environmental bacteria. The phytochemical berberine inhibits the growth of *E. coli* (Sun *et al.*, 2019b), and berberine chloride exposure leads to upregulation of the RND pump *acrF* (Karaosmanoglu *et al.*, 2014). Deletion of *tolC* also leads to a significant increase in berberine susceptibility (Tegos *et al.*, 2002).

Efflux pumps also play a vital role in maintaining metal ion homeostasis, CusABC in *E. coli* is responsible for controlling both copper and silver ion levels (Grass and Rensing, 2001, Franke *et al.*, 2001). Furthermore, heavy metals have been shown to increase transcription of efflux pump components in *Pseudomonas aeruginosa* (Martinez *et al.*, 2009).

Additionally, in some species efflux pumps are involved in cell-to-cell communication. Quorum sensing (QS) is a density dependent response in which signalling molecules called autoinducers are released into the microenvironment. When the concentration of these autoinducing molecules reaches a certain level, they bind to an associated autoinducer regulator, resulting in expression of a set of genes. Functions of QS-regulated genes include antimicrobial peptide synthesis and motility (Martinez *et al.*, 2009). The RND pump BpeAB-OprB in *Burkholderia pseudomallei* was shown to efflux five different homoserine lactone autoinducers involved in QS (Chan *et al.*, 2007).

1.12 Efflux inhibition

With the continuing increase in MDR infections, new therapeutic alternatives are needed. The vast majority of clinically used antimicrobials were discovered and developed during the so called ‘Golden Era’ between 1941 and 1968 (Li and Nikaido,

2009). Since then few new and novel compounds have been developed. Efflux inhibitors (EI) are therefore an attractive solution that may enable the potentiation of ineffective antimicrobials, while also potentially having anti-virulence and anti-biofilm properties (Reza *et al.*, 2019). Despite extensive research into EI design and function, no EI has yet made it to clinic.

EIs can be competitive by which the inhibitor is also a substrate of the efflux system and therefore competes without other substrates. In the case of antimicrobials, this would slow the rate by which compounds are removed from the cell, increasing the time for potential interaction with the molecule target. An example of a competitive inhibitor is phenylalanine-arginine β-naphthylamide (PAβN). Non-competitive EIs work indirectly and do not compete with substrate binding but instead inhibit efflux via alternative methods e.g. preventing complex assembly or dissipating the energy source as is the case for carbonyl cyanide m-chlorophenyl hydrazone (CCCP), which depolarises the cell membrane and disrupts the proton motive force, thereby rendering efflux pumps (excluding ABC pumps) non-functional (Sharma *et al.*, 2019).

While no EI has made it to clinic to date, the best characterised RND EI is PAβN which is a substrate of AcrB and was found to potentiate fluoroquinolones, chloramphenicol, rifampicin and macrolides, reduce biofilm produce and the expression virulence genes (Lomovskaya and Bostian, 2006, Matsumoto *et al.*, 2011, Kvist *et al.*, 2008, Rampioni *et al.*, 2017). However, essential moieties in PAβN and subsequent derivatives have cytotoxic activity (Lomovskaya and Bostian, 2006). Furthermore, the compound has several off target effects including membrane permeabilisation (Lamers *et al.*, 2013). AcrB specific inhibitors, such as MBX 2319 and

MBX 3135, that posess activity at significantly lower concentrations than PAβN, have since been described (Sjuts *et al.*, 2016).

Another potential approach for inhibition is the use of natural products. The *K. pneumoniae* bacteriocin, klebcin has been shown to enter the cell via the OMP TolC. Removal of the TonB motif required for its translocation results in the toxin binding within the TolC lumen effectively blocking the channel (Housden *et al.*, 2021). Klebcin was shown to both inhibit the efflux of the TolC substrate Nile red and growth in the presence of rhodamine 6G (Housden *et al.*, 2021).

1.13 Sequence variation and/or conservation across RND efflux pumps

Much of what we know about efflux pump structure and function, is based on the pump sequences found in type strains of Gram-negative bacterial species such as *E. coli* MG1655 or *Pseudomonas aeruginosa* PA01. Furthermore, efforts to develop EIs also concentrate on these laboratory strains. Though useful for elucidating fundamental processes, laboratory strains do not represent the true genetic diversity across a species (Hobman *et al.*, 2007).

E. coli has six RND pump systems, whereas *Salmonella* is reported to have five (Anes *et al.*, 2015, Blair *et al.*, 2015b). Environmental bacteria like *E. coli* and *P. aeruginosa* have significantly more RND pump systems than obligate human pathogens such as *Chlamydia trachomatis* and *Haemophilus influenzae,* which have adapted to a specific host and environment (Nikaido and Zgurskaya, 2001) suggesting a correlation between lifestyle and RND number.

Phylogenetic analysis of RND pump sequences identified seven families within the RND superfamily (Tseng *et al.*, 1999). A similar phylogenetic analysis of 236 PAPs

(one representative sequence per species) identified eight main sequence clusters and also identified highly conserved residues across all species. The most conserved region of the protein was found to be adjacent to the β-barrel domain (Zgurskaya *et al.*, 2009). While both studies identified key residues, neither looked at the extent of conservation across a species, instead relying on representative protein sequences.

AcrAB-TolC is found in multiple species of the Enterobacteriaceae but gene sequences and subsequent amino acid sequences vary e.g. the amino acid similarity of AcrB in *K. pneumoniae, E. cloacae,* and *S.* Typhimurium to *E. coli* is 96%, 96%, and 97%, respectively. Furthermore, single nucleotide polymorphisms (SNPs) have been described within the RND efflux pumps of clinical isolates. SNPs are likely to be silent and therefore do not result in an amino acid change, however some will confer a phenotypic advantage over competing cells. A Q733R substitution in AcrB modestly contributed to the resistance profile of *E. coli* strains, while a G288D substitution was found to confer resistance to clinically relevant drugs (Elkins *et al.*, 2010, Blair *et al.*, 2015a). A similar substitution at residue 714 in the single RND pump of *Neisseria gonorrhoeae,* MtrD, has been reported in clinical isolates resulting in reduced macrolide susceptibility (Ma *et al.*, 2020). Elkins *et al.* (2013) reported that isolates containing non-silent SNPs were more associated with environments previously exposed to antimicrobials.

1.14 Aims of work

Fundamental biology provides the foundation for rational design of novel antimicrobials and EIs. While our understanding of RND structures and their interactions with PAPs and OMPs is expanding, it is not clear how representative laboratory strains, such as

E. coli K-12 and *S.* Typhimurium SL1344, are of clinical isolates. This project set out to determine the conservation of RND efflux pumps and their associated components across the diversity of *E. coli.*

Chapter 2 Materials and methods

2.1 Basic microbiology

The wild-type strain and background used for genetic manipulation in Chapters 1, 2 and 3 was *Escherichia coli* K-12 MG1655 (Blattner *et al.*, 1997). In Chapter 4, both *E. coli* K-12 MG1655 and *Escherichia coli* ATCC 25922 (PHE) were used (Table 2.1). Bacterial strains were stored at -80˚C using Protect cryopreservation beads (Technical Service Consultants). When required for experiments, a bead of the desired strain was spread onto Luria Bertani (LB) agar (Sigma-Aldrich; L2897) and incubated overnight at 37˚C unless stated otherwise. Subsequent liquid cultures were then prepared from this plate with 5 mL of LB broth (Sigma-Aldrich; L3022) inoculated with a single colony of bacteria. The broth was then incubated at 37˚C with aeration.

Several plasmids were used throughout the course of this work for genetic manipulation of strains and cloning (Table 2.2). Plasmids were maintained in strains by supplementing LB with the antibiotic required. Both ampicillin and kanamycin were used in this study at a final concentration of 50 μg/mL.

Table 2.1 Strains used in the study

Laboratory code	Species	Genotype	Antibiotic resistance	Reference
EC07	*E. coli*	CGSC - DB Strain K-12 MG1655		(Blattner *et al.*, 1997)
EC15	*E. coli*	K-12 MG1655 Δ*acrA*		Blair Group strain collection
EC98	*E. coli*	K-12 MG1655 Δ*acrA* + pET21a	Amp^R	This study
EC65	*E. coli*	K-12 MG1655 Δ*acrA* + pET21a::*acrA*	Amp^R	This study
EC78	*E. coli*	K-12 MG1655 Δ*acrA* + pET21a::*acrA* T104A	Amp^R	This study
EC29	*E. coli*	K-12 MG1655 Δ*acrB*		(Wang-Kan *et al.*, 2017)
EC99	*E. coli*	K-12 MG1655 Δ*acrB* + pET21a	Amp^R	This study
EC66	*E. coli*	K-12 MG1655 Δ*acrB* + pET21a::*acrB*	Amp^R	This study
EC84	*E. coli*	K-12 MG1655 Δ*acrB* + pET21a::*acrB* H596N	Amp^R	This study

EC62	*E. coli*	K-12 MG1655 Δ*tolC*		Blair Group strain collection
EC100	*E. coli*	K-12 MG1655 Δ*tolC* + pET21a	Amp^R	This study
EC67	*E. coli*	K-12 MG1655 Δ*tolC* + pET21a::*tolC*	Amp^R	This study
EC68	*E. coli*	K-12 MG1655 Δ*tolC* + pET21a::*tolC* A233T	Amp^R	Blair Group strain collection
EC102	*E. coli*	NCTC 12900 O157:H7 (Shiga toxin negative)		PHE
EC115	*E. coli*	K-12 MG1655 *acrF*::*aph*	Kan^R	This study
EC116	*E. coli*	K-12 MG1655 Δ*acrF*		This study
EC117	*E. coli*	K-12 MG1655 Δ*acrB acrF*::*aph*	Kan^R	This study
EC118	*E. coli*	K-12 MG1655 Δ*acrB* Δ*acrF*		This study
EC232	*E. coli*	K-12 MG1655 Δ*acrF* + pET21a	Amp^R	This study
EC233	*E. coli*	K-12 MG1655 Δ*acrB* Δ*acrF* + pET21a	Amp^R	This study

EC327	*E. coli*	K-12 MG1655 Δ*acrB* Δ*acrF* + pET21a::*acrF* (K12 sequence)	Amp^R	This study
EC328	*E. coli*	K-12 MG1655 Δ*acrB* Δ*acrF* + pET21a::*acrF* (ST11 sequence)	Amp^R	This study
EC112	*E. coli*	ATCC 25922		PHE
EC177	*E. coli*	ATCC 25922 *eefB*::*aph*	Kan^R	This study
EC188	*E. coli*	ATCC 25922 Δ*eefB*		This study
EC195	*E. coli*	ATCC 25922 *eefABC*::*aph*	Kan^R	This study
EC196	*E. coli*	ATCC 25922 Δ*eefABC*		This study
EC210	*E. coli*	ATCC 25922 *eefD*::*aph*	Kan^R	This study
EC227	*E. coli*	ATCC 25922 Δ*eefD*		This study
EC218	*E. coli*	K-12 MG1655 pACYC177	Amp^R Kan^R	This study
EC201	*E. coli*	K-12 MG1655 + pACYC177 *eefD*	Kan^R	This study

EC219	*E. coli*	K-12 MG1655 Δ*acrB* pACYC177	Amp^R Kan^R	This study
EC220	*E. coli*	K-12 MG1655 Δ*acrB* pACYC177 *eefD*	Kan^R	This study
EC97	*E. coli*	K-12 MG1655 + pET21a	Amp^R	This study
EC378	*E. coli*	K-12 MG1655 + pET24a	Kan^R	This study
EC221	*E. coli*	K-12 MG1655 Δ*acrB* pET24a	Kan^R	This study
EC369	*E. coli*	K-12 MG1655 + pET21a::*eefABC*	Amp^R	This study
EC370	*E. coli*	K-12 MG1655 + pET24a::*eefABC*	Kan^R	This study
EC371	*E. coli*	K-12 MG1655 Δ*acrB* + pET21a::*eefABC*	Amp^R	This study
EC372	*E. coli*	K-12 MG1655 Δ*acrB* + pET24a::*eefABC*	Kan^R	This study
EC10	*E. coli*	BW1125 + pKD4 Datsenko & Wanner	Amp^R	(Datsenko and Wanner, 2000)
EC23	*E. coli*	K-12 MG1655 + pSLTS	Amp^R	This study

EC164	*E. coli*	*E. coli* 1200 + pACYC177	Amp^R Kan^R	(Chang and Cohen, 1978)
EC113	*E. coli*	K-12 BW25113 *acrF::aph* (Keio Collection strain; 63:6B)	Kan^R	(Baba *et al.*, 2006)
SE44	*S.* Typhimurium	14028s		ATCC
SE160	*S.* Typhimurium	SL1344 + pCP20	Amp^R	(Cherepanov and Wackernagel, 1995)

Table 2.2 Plasmids used in this study

Plasmid	Description	Antibiotic resistance marker	Source
pKD4	Encodes *aph* resistance gene flanked by FRT sites.	Kan^R, Amp^R	(Datsenko and Wanner, 2000)
pSLTS	Encodes an inducible λ-red recombinase I-SceI.	Amp^R	(Kim *et al.*, 2014)
pET21a	High copy cloning vector with *bla* resistance gene. Possesses a pBR322 origin of replication.	Amp^R	Novagen/Thermo Fisher
pET24a	High copy cloning vector with *aph* resistance gene. Possesses a pBR322 origin of replication.	Kan^R	Novagen/Thermo Fisher
pACYC177	Low copy cloning vector with p15A origin of replication.	Kan^R, Amp^R	(Chang and Cohen, 1978)
pCP20	Encodes the inducible Flp recombinase.	Amp^R	(Cherepanov and Wackernagel, 1995)

2.2 Isolation of plasmid DNA

A 10 mL culture of the plasmid carrying strain was grown overnight at 37˚C (30˚C if temperature sensitive) with aeration. LB was supplemented with the required antibiotic for plasmid maintenance. Cultures were pelleted at 1523 x *g* for 10 minutes, the pellet was then used for plasmid isolation. The Qiagen QIAprep Spin Miniprep Kit (Qiagen; Cat. No. 27106) was used as described in the manufacturer's instructions. Plasmid elutions were made using 30 µL UltraPure DNase/RNase-free distilled water and stored at -20˚C.

2.3 Production of electrocompetent cells

A single colony was inoculated in 5 mL LB broth (supplemented with antibiotics if necessary) and grown overnight at 37˚C with aeration. The following morning 2 mL of this culture was used to inoculate 50 mL of LB in a 500 mL Duran bottle to ensure oxygen availability. Cells were grown at 37˚C with aeration until an optical density wavelength at 600 nm (OD_{600}) of 0.6 was achieved. This was measured using Jenway 6300 spectrometer (Cole-Parmer Ltd). Thereafter cells were maintained on ice.

Cells were transferred into a 50 mL Falcon tube and harvested by centrifugation at 2986 x *g* for 10 minutes at 4˚C. Supernatant was then removed, and cells resuspended in 25 mL ice-cold glycerol (15% v/v). Cells were washed two more times and the final pellet was resuspended in 500 µL of glycerol. This was then aliquoted into reaction tubes, each containing 50 µL of cells ready for electroporation and stored at -20˚C.

2.4 Bacterial transformation

Aliquots of prepared electrocompetent cells were transferred into a 2 mm electroporation cuvette (Geneflow, UK; Cat. No. E6-006) and incubated on ice for 5 minutes. A range of 1 and 5 µL of prepared DNA was added to the cuvette which was gently tapped to encourage mixing. Electroporation was performed at 2500 V (25 µF, 200 Ω) using an Eppendorf Eporator (Eppendorf, UK). As a negative control cells not mixed with any vector were also electroporated. Following the electrical impulse, cells were immediately resuspended in 950 µL of warm LB broth and transferred into a sterile universal. Cells were then recovered at the appropriate temperature for an hour with aeration.

Following recovery, 50 µL of cells were then spread onto prewarmed LB agar supplemented with the required antibiotic. Remaining cells were pelleted at 18407 x *g* for 3 minutes, resuspended in LB broth and spread onto an additional LB agar plate supplemented with the required antibiotic. LB broth used for cell recovery was also spread on LB and supplemented LB as a further sterility control.

2.5 PCR amplification conditions

DNA of interest was amplified using an Eppendorf Nexus X2 thermal cycler (Eppendorf, UK). DNA fragments for cloning and sequencing were generated using Q5 Hot Start High Fidelity 2x Master Mix (New England BioLabs; M0494S). General PCRs were performed using MyTaq 2x Red (Meridian; Cat. No. BIO21109). PCR mixtures contained 12.5 µL enzyme, 1 µL of 10 µM forward primer, 1 µL 10 µM reverse primer, 1 µL of template DNA and were then made up to 25 µL using nuclease free water. Oligonucleotides used in this study are listed in Table 2.3 and were obtained

from Sigma-Aldrich (Merck Life Science, UK) or Eurogentec (Kaneka Eurogentec S.A., Belgium).

Both MyTaq and Q5 enzymes have an extension rate of 1 kilobase of DNA per 30 seconds. All extensions were therefore calculated with this rate in mind and the general PCR programme set-up can be found in Table 2.4.

Table 2.3 Primers used in this study

Laboratory code	Primer name	Primer description	Primer sequence 5'-3'
P119	EC *acrA* internal Fwd	*Escherichia coli* MG1655 internal *acrA* check Fwd	AGAGGGTTTACGCCTCTGG
P120	EC *acrA* internal Rev	*Escherichia coli* MG1655 internal *acrA* check Rev	TGACCGTTCTGTACCAATGC
P121	EC *acrB* internal Fwd	*Escherichia coli* MG1655 internal *acrB* check Fwd	AATGGCGCTGTCGGTACT
P122	EC *acrB* internal Rev	*Escherichia coli* MG1655 internal *acrB* check Rev	CCCAGAGTGGTGTTAATGTCG
P123	EC *tolC* internal Fwd	*Escherichia coli* MG1655 internal *tolC* check Fwd	ACACAGGCACAAAAAGAAGC
P129	pET21a Fwd	pET21a primer Fwd (T7 promoter primer - 69348-3)	TAATACGACTCACTATAGGG
P130	pET21a Rev	pET21a primer Rev (T7 Terminator Primer - 69337-3)	GCTAGTTATTGCTCAGCGG
P131	EC *tolC* check F	*Escherichia coli* MG1655 external *tolC* check Fwd	ACGTAACGCCAACCTTTTGC
P132	EC *tolC* check R	*Escherichia coli* MG1655 external *tolC* check Rev	TTTGTCTTCCGGGACCAGTG
P133	EC *acrB* check F	*Escherichia coli* MG1655 external *acrB* check Fwd	CTGATCACCAGTGACGGCAT
P134	EC *acrB* check R	*Escherichia coli* MG1655 external *acrB* check Rev	CGTATGAGATCCTGAGTTGG

P262	*acrF* check F	Forward primer for external check of *acrF* in *E. coli* MG1655	attggcgataaatggttgat
P263	*acrF* check R	Reverse primer for external check of *acrF* in *E. coli* MG1655	tcttttcatggtgttacctc
P275	EC *acrF* cloning *Xho*I R	Reverse primer for cloning *acrF* from EC07 and EC102 into the pET21a vector. *Xho*I restriction site downstream of internal stop codon.	ggtgctcgagTTATCCTTTAAAGCAACGGC
P290	EC *eefABC* cloning *Xho*I R extended	Extended homologous region reverse primer for cloning *eefABC* from *E. coli* ATCC 25922 onto pET21a vector. Incorporates *Xho*I site. Product is on complement strand.	GTGGTGCTCGAGTTACTTGAGTGATGACATACCGCCG
P301	EC *acrF* internal check F	Forward primer for amplifying a 1521 bp internal region of *acrF* in *E. coli* MG1655	GAAGAATTCGGCAAAGTGAC
P302	EC *acrF* internal check R	Reverse primer for amplifying a 1521 bp internal region of *acrF* in *E. coli* MG1655	TGAAATGGTCTGATTGATGT
P303	EC *acrF* internal check O157:H7 F	Forward primer for amplifying a 1270 bp internal region of *acrF* in *E. coli* O157:H7 - binds to insertion	TTCACATAACTGTTGCGT
P304	EC *acrF* internal check O157:H7 R	Reverse primer for amplifying a 1270 bp internal region of *acrF* in *E. coli* O157:H7 - binds to insertion	ATCCTTTAAAGCAACGGC
P311	EC *acrF* primer fragment 1 R	Reverse primer, used in combination with P129, amplifies fragment 1 of *acrF* from pET vector for sequencing.	GGGAAGAATGGCTGTAATTC

P312	EC *acrF* primer fragment 4 F	Reverse primer, used in combination with P313, amplifies fragment 4 of *acrF* from pET vector for sequencing.	ATGATGCGATAGTGGTGGTG
P313	EC *acrF* primer fragment 4 R	Reverse primer, used in combination with P312, amplifies fragment 4 of *acrF* from pET vector for sequencing.	TTGGTGCTGCGGACATAAAG
P314	EC *acrF* primer fragment 5 F	Reverse primer, used in combination with P130, amplifies fragment 5 of *acrF* from pET vector for sequencing.	TGAAACCGTGGGAAGAGC
P394	EC *eefB* KO primer forward	Forward primer to introduce *aph* cassette in *eefB* in *E. coli* ATCC 25922	AGGCTGTTGAAGTTCAGGCTAATGGAGGCAACGCCTGATGTTTgtgtaggctggagctgcttc
P395	EC *eefB* KO primer reverse	Reverse primer to introduce *aph* cassette in *eefB* in *E. coli* ATCC 25922	ACAACACCAGAAAAATAAGAGAACGACGCAACATAACTTAgggaattagccatggtccat
P396	EC *eefA-eefC* KO primer forward	Forward primer to introduce *aph* cassette in *eef* operon in ATCC 25922	TTCAATGAGGAGGGTATTTCTCACCCTGGAGAAAATAATGgtgtaggctggagctgcttc
P397	EC *eefA-eefC* KO primer reverse	Reverse primer to introduce *aph* cassette in *eef* operon in ATCC 25922	AATGCCCATGAAAGAGAGACTCTAGCCATATTATTCCTTAgggaattagccatggtccat
P409	*eefD Apa*LI clone F	Forward primer to clone *eefD* from ATCC 25922 into pACYC177; has *Apa*LI site (uses plasmid's native Amp promoter)	CCCCCCGTGCACATGGCTAGAGTCTCTCTTTCATGGGCATTGATTCTTGG
P410	*eefD Pst*I clone R	Reverse primer to clone *eefD* from ATCC 25922 into pACYC177; has *Pst*I site (uses plasmid's native Amp promoter)	CCCCCCCTGCAGTTACTTTTGATTGTCGATTCTGGTGAGAGCAAC

P411	pACYC177 Amp F chk	Forward primer to check the product size of the AmpR region in pACYC177 (*eefD* = 1872 bp, *ampR* = 1128 bp)	GTTTATAAAATTCTTGAAGACG
P412	pACYC177 Amp R chk	Reverse primer to check the product size of the AmpR region in pACYC177 (*eefD* = 1872 bp, *ampR* = 1128 bp)	GAAGTTTTAAATCAATCTAAAG
P413	*eefB* KO chk F	Forward primer to confirm the size of the *eefB* gene in ATCC 25922	GGTGAATAAAGACAATAAAG
P414	*eefB* KO chk R	Reverse primer to confirm the size of the *eefB* gene in ATCC 25922	CTGAGTCAATATCAGCAATC
P415	*eefABC* KO chk F	Forward primer to confirm the size of the *eefABC* operon in ATCC 25922	CATTCGTTGCTGCGTTGAAG
P416	*eefABC* KO chk R	Reverse primer to confirm the size of the *eefABC* operon in ATCC 25922	GGTTAATTGCGTTATGGTCG
P472	pET24a *aph* chk F (Pharmacia)	*aph* check for pET24a	gagaaaactcaccgaggcag
P473	pET24a *aph* chk R (Pharmacia)	*aph* check for pET24a	gtatttcgtctcgctcaggc
P476	EC *eef* frag 1 R	To be used with P130 for amplification of *eef* operon for sequencing; fragment 1	TAA CGC TGA TTG CTG AAA TC
P477	EC *eef* frag 2 F	Amplification of *eef* operon for sequencing; fragment 2	TGT TGC TGG CGC TGG ATA AT
P478	EC *eef* frag 2 R	Amplification of *eef* operon for sequencing; fragment 2	TGT GGA ATT TGC TGA AGC GG
P479	EC *eef* frag 3 F	Amplification of *eef* operon for sequencing; fragment 3	GCC AGC GAG GTC ATG ATG AT

P480	EC *eef* frag 3 R	Amplification of *eef* operon for sequencing; fragment 3	CAA ATG CCG CAA TTG CAG GT
P481	EC *eef* frag 4 F	Amplification of *eef* operon for sequencing; fragment 4	CAC TGG ACA GGG TGT CGG TG
P482	EC *eef* frag 4 R	Amplification of *eef* operon for sequencing; fragment 4	GCA CAT TCG GCG TTC TTG CG
P483	EC *eef* frag 5 F	Amplification of *eef* operon for sequencing; fragment 5	CTA CCA CCA CAA TGG CAT CG
P484	EC *eef* frag 5 R	Amplification of *eef* operon for sequencing; fragment 5	TGA CAC CAC TGA CAA AGC CT
P485	EC *eef* frag 6 F	Amplification of *eef* operon for sequencing; fragment 6	AGA CTC CCC ACA CCT TCA AC
P486	EC *eef* frag 6 R	Amplification of *eef* operon for sequencing; fragment 6	TGA CAT TAC GGG CAA TTT TC
P487	EC *eef* frag 7 F	To be used with P129 for amplification of *eef* operon for sequencing; fragment 7	ACT TTG GCA GAA CCT TC
P488	EC *eefD* KO primer forward	Primer to create *eefD* inactivation R	CTGGGCGGCGGTATGTCATCACTCAAGTAAGGAATAATgtgtaggctggagctgcttc
P489	EC *eefD* KO primer reverse	Primer to create *eefD* inactivation F	GCACCTCATCCTGACAGGCCCCCTCAGGATGAGGTGTgggaattagccatggtccat
P533	EC *eefD* KO check F	Forward primer to confirm length of *eefD* in ATCC 25922 (EC112) [doesn't work on delta *eefABC*]	GTCGTCGTTGTTCAGCCACG
P534	EC *eefD* KO check R	Reverse primer to confirm length of *eefD* in ATCC 25922 (EC112) [doesn't work on delta *eefABC*]	TATTGCCGATGAAGCCCTCG

P540	EC *eefD* KO check F long	Forward primer to confirm length of *eefD* in ATCC 25922 (EC112); extended to try to correct for nonspecific bands [doesn't work on delta *eefABC*]	CGCATTGTCGTCGTTGTTCAGCCACGGAAT
P541	EC *eefD* KO check R long	Reverse primer to confirm length of *eefD* in ATCC 25922 (EC112); extended to try to correct for nonspecific bands [doesn't work on delta *eefABC*]	TGCGCCCAAAAGTCATTATTGCCGATGAAG
P650	pET MCS checker	Primer to confirm which MCS was used in pET21/24a. Use with P130	gaaacaagcgctcatgag
P653	pET21/24 *acrF Nde*I F	Works with P275 to clone *acrF* into pET21a or pET24a. Can be used on normal *acrF* and O157 sequence. Incorporates the *Nde*I site upstream of start site.	ttgaccatatgGCAAACTTTTTTATTCGAC
P656	pET21/24 *eefABC Nde*I F	Works with P290 to clone *eefABC* into either pET21 and pET24. Incorporates the *Nde*I site upstream of start site.	ttgaccatATGAAGTATATAGCAACATCTGTAGTG

Table 2.4 Generic PCR conditions

Stage	Time	Temperature	Number of cycles
Initial denaturation	5 minutes	95˚C	1
Denaturation	15 seconds	95˚C	30
Annealing	15 seconds	*˚C	
Extension	30 seconds per kb of amplimer	72˚C	
Final extension	10 minutes	72˚C	1
Hold	Indefinite	4˚C	1

*˚C – annealing temperature was dependent on the predicted Tm values for each primer set.

2.6 DNA purification

PCR products were separated by gel electrophoresis at 80 V on a 1% (w/v) agarose gel. The gel was prepared by dissolving agarose (Sigma; A9539) in 100 mL 1x tris-acetate-EDTA buffer (Sigma; Cat. No. T9650) supplemented with 5 µL Midori Green (Nippon Genetics Europe; MG 03). Along with PCR products, 5 µL 1 kb Hyperladder (Bioline; Cat. No. BIO33025) and a negative control were electrophoresed. Gels were imaged using an Amersham Imager 680 (General Electronics, USA) under UV-light (312 nm).

If multiple bands were present on the image, the remaining PCR products were electrophoresed and amplimers of interest excised under blue light (Illuminator; Nippon Genetics Europe). Gel products were then purified using the QIAquick Gel Extraction Kit (Qiagen; Cat. No. 28706) following manufacturer's instructions. DNA was eluted in 30 µL UltraPure DNase/RNase-free distilled water and DNA stored at -20˚C.

If samples contained only the DNA amplimer, the remaining PCR product was purified using the QIAquick PCR Purification Kit (Qiagen; Cat. No. 28106) following

manufacturer's instructions. DNA was eluted in 30 µL UltraPure DNase/RNase-free distilled water and DNA was stored at -20˚C.

DNA concentration was determined using a Nanodrop spectrophotometer 2000 (Thermo Scientific).

2.7 DNA sequencing of PCR products

Sanger sequencing performed by Eurofins Genomics (Germany) was used to confirm PCR amplimers. Purified PCR products were diluted to a final concentration of 10 ng in 15 µL UltraPure DNase/RNase-free distilled water and mixed with 2 µL of 10 µM forward or reverse primer. Sequencing results were then confirmed using alignment using EBI Clustal Omega (Madeira *et al.*, 2019).

2.8 Gene cloning

2.8.1 Cloning of *acrF* from K-12 and ST11 into pET21a

The *acrF* genes from *E. coli* K-12 MG1655 (EC07) and *E. coli* NCTC 12900 (ST11) (EC102) were amplified using primers that introduced the *Nde*I and *Xho*I restriction sites (Table 2.3). Reactions contained 12.5 µL Q5 Hot Start High Fidelity 2x Master Mix (New England BioLabs), 1 µL P275 (10 µM), 1 µL P653 (10 µM), 1 µL dimethyl sulfoxide (DMSO), 1.5 µL DNA (cell lysate) and made up to 25 µL using UltraPure DNase/RNase-free distilled water. The resulting PCR products were purified following gel electrophoresis and elutions stored at -20˚C.

The pET21a plasmid (Thermo Fisher) was purified from EC97 (*E. coli* K-12 + pET21a). The plasmid and *acrF* fragments from K-12 and ST11 were then digested using the restriction enzymes *Nde*I (New England BioLabs; Cat. No. R0111S) and *Xho*I (New England BioLabs; Cat. No. R0146S). The reaction set-up consisted of 1 µg DNA,

1 µL of each restriction enzyme, 5 µL of 10x reaction buffer, and UltraPure DNase/RNase-free distilled water up to 50 µL. Undigested and single enzyme controls were also performed to confirm restriction enzyme activity. Reactions were incubated at 37˚C for 1 hour. After 1 hour, 1 µL calf intestinal phosphatase (New England BioLabs; Cat. No. M0290S) was added to the double-cut plasmid reaction, and this reaction was incubated for a further 1 hour at 37˚C.

Following digestion, the double cut *acrF* gene fragments were purified using the QIAquick PCR Purification Kit (Qiagen; Cat. No. 28106). Uncut, single- and double-cut pET21a reactions were electrophoresed on a 1% (w/v) agarose gel. Following separation, the single cut and double cut plasmids were excised from the gel with a scalpel under blue light. The single cut bands were stored at -20˚C and double cut band purified using the QIAquick Gel Extraction Kit (Qiagen; Cat. No. 28706). DNA samples were eluted in 30 µL UltraPure DNase/RNase-free distilled water.

Digested *acrF* inserts and the double cut pET21a vector were ligated at a 3:1 molar ratio using T4 ligase (New England BioLabs; M0202S). Ligation reactions consisted of 1 µL T4 ligase, 2 µL 10x T4 ligase buffer, 3:1 ratio of insert to vector, and UltraPure DNase/RNase-free distilled water up to a final volume of 20 µL. A background control of double cut pET21a without insert was also prepared. Reactions were incubated at 16˚C overnight. The following morning, the T4 ligase was heat inactivated by incubation at 65˚C for 10 minutes. If ligation reactions were not transformed into a background strain that day, they were stored at -20˚C. Successful ligations were confirmed via PCR and sequencing as described in 2.7 (Figure 2.1 and Figure 2.2).

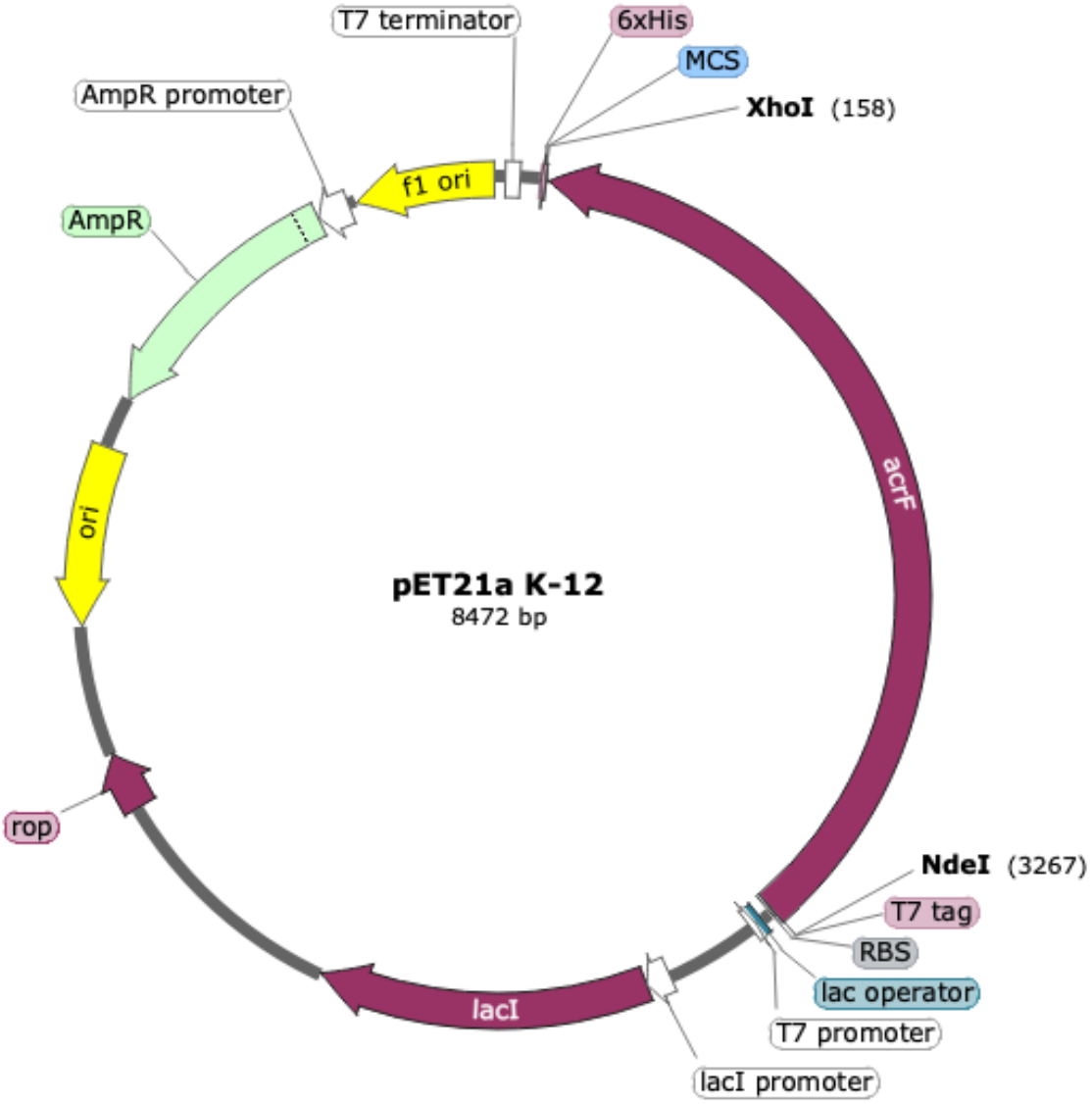


Figure 2.1 Plasmid map for pET21 *acrF* K-12

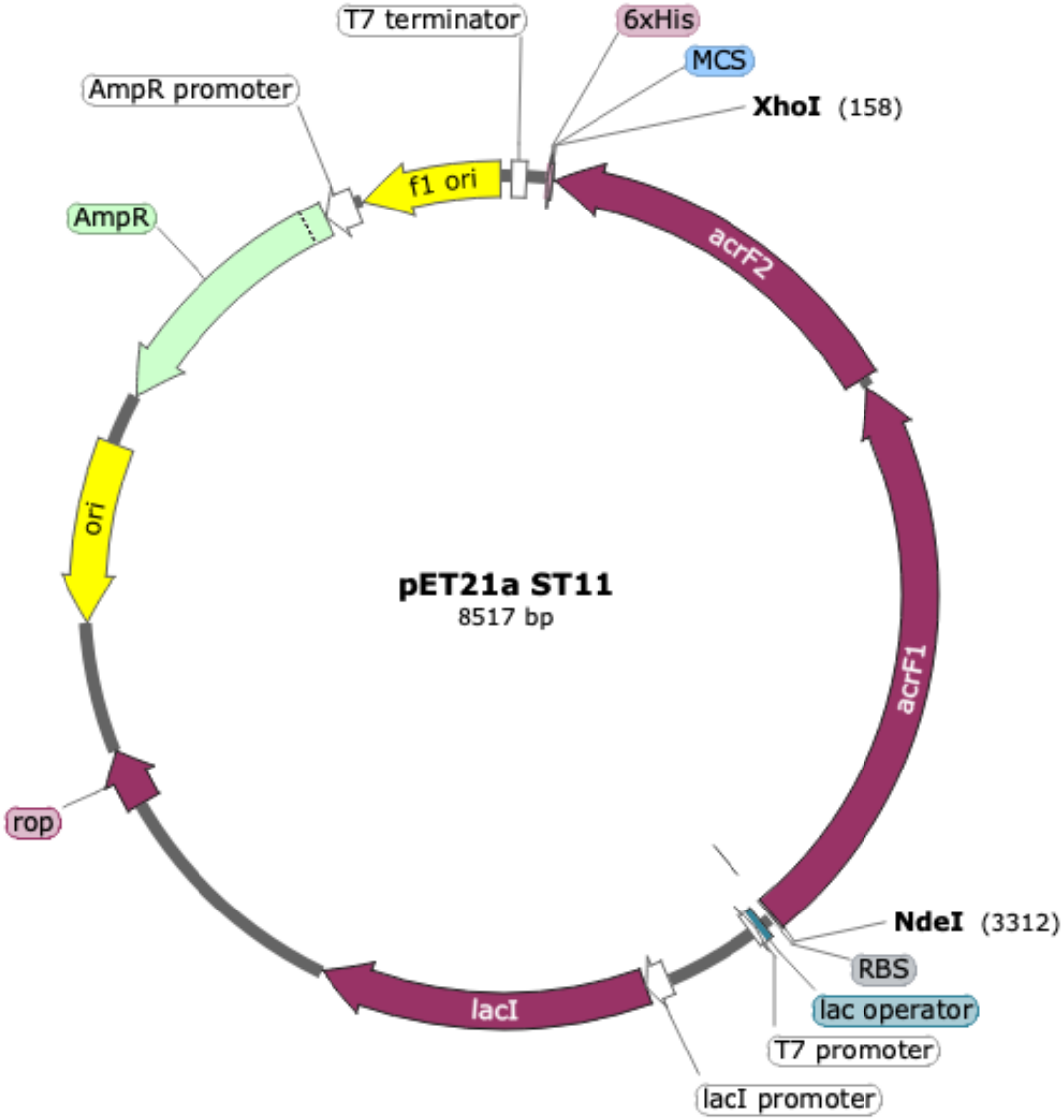


Figure 2.2 Plasmid map for pET21a *acrF* ST11

2.8.2 Cloning of *eefABC* into pET21a and pET24a

eefABC was amplified from *E. coli* ATCC 25922 (EC112) with primers P290 and P656 which incorporate the *Nde*I and *Xho*I restrictions sites. The amplification reaction contained 12.5 µL Q5 Hot Start High Fidelity 2x Master Mix, 1 µL P290 (10 µM), 1 µL P656 (10 µM), 1.5 µL DNA (colony lysate), 1 µL DMSO, made up to 25 µL with UltraPure DNase/RNase-free distilled water. PCR products were separated via gel electrophoresis and purified using the QIAquick Gel Extraction Kit. Purified DNA was eluted in 30 µL UltraPure DNase/RNase-free distilled water. To aid elution, water was heated to 60˚C prior to eluting.

The pET21a plasmid was isolated from EC97 and pET24a from EC189. The *eefABC* DNA fragment and both pET21a and pET24a were digested using the *Nde*I and *Xho*I restriction enzymes; pET21a and pET24a have the same multiple cloning site and only differ by the antibiotic resistance marker (Table 2.2). The restriction digestion and ligation were performed as described in 2.8.1. Successful ligations were confirmed via PCR and sequencing as described in 2.7 (Figure 2.3 and Figure 2.4).



Figure 2.3 Plasmid map for pET21a *eefABC*

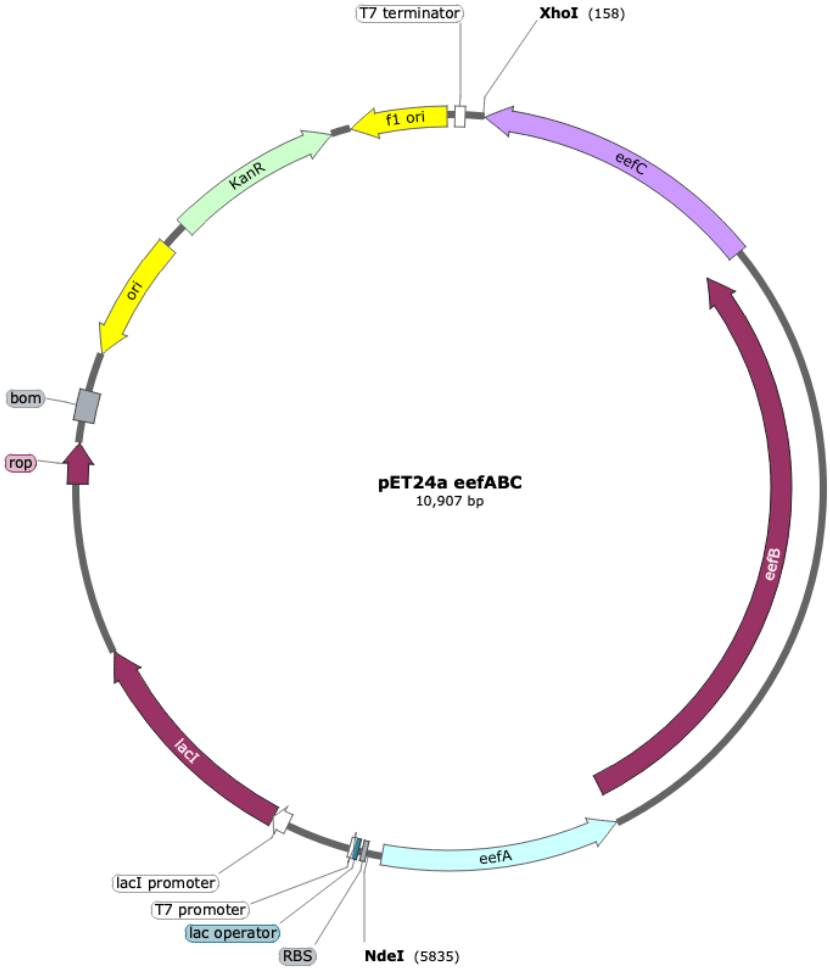


Figure 2.4 Plasmid map for pET24a *eefABC*

2.8.3 Cloning of *eefD* into pACYC177

The *eefD* gene, encoding an MFS pump in *E. coli* ATCC 25922 EC112, was amplified by PCR using primers P409 and P410 which incorporated the *Apa*LI and *Pst*I restriction sites. The PCR reaction was as follows; 12.5 µL Q5 Hot Start High Fidelity 2x Master Mix, 1 µL P409 (10 µM), 1 µL P410 (10 µM), 1 µL DMSO, 1.5 µL DNA (colony lysate), made up to 25 µL with UltraPure DNase/RNase-free distilled water. PCR products were electrophoresed, purified and eluted in 30 µL UltraPure DNase/RNase-free distilled water.

The pAYCC177 plasmid was isolated from EC164 using the QIAprep Spin Miniprep Kit following manufacturer's instructions and eluted in 30 µL UltraPure DNase/RNase-free distilled water. Both the plasmid and the *eefD* amplimer were digested using *Apa*LI and *Pst*I in the same way as described in 2.8.1. Briefly the reaction consisted of 1 µg DNA, 1 µL of each restriction enzyme, 5 µL of 10x reaction buffer, and UltraPure DNase/RNase-free distilled water up to 50 µL. Uncut and single cut reactions were also included. All reactions were incubated at 37˚C for 1 hour. Double cut plasmid was also treated with calf intestinal phosphatase for 1 hour at 37˚C. The digested *eefD* amplimer was purified using the QIAquick PCR purification kit, while uncut, single- and double-cut plasmids were separated via gel electrophoresis and then purified using the QIAquick Gel Purification Kit. All purified DNA was eluted in 30 µL UltraPure DNase/Rnase-free distilled water.

Ligation of pACYC177 and *eefD* fragments was achieved using T4 ligase as described in 2.8.1. Successful ligations were confirmed via PCR and sequencing as described in 2.7 (Figure 2.5).

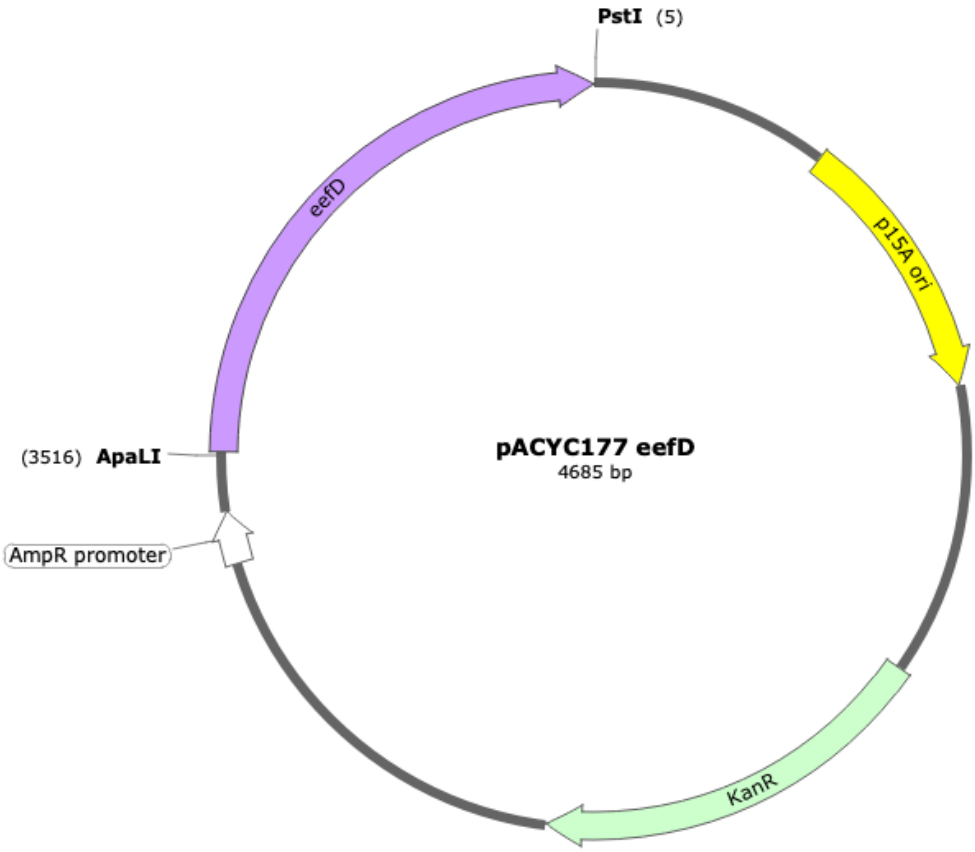


Figure 2.5 Plasmid map for pACYC177 *eefD*

2.9 Gene disruption in *E. coli*

2.9.1 Homologous recombination

Genes of interest were disrupted using an adapted version of the protocol described by Datsenko and Wanner (Kim *et al.*, 2014, Datsenko and Wanner, 2000). Briefly, this method uses homologous recombination to introduce an aminoglycoside-3'-phosphotransferase gene (*aph*) in place of the gene of interest.

2.9.1.1 Generation of *aph* cassette

This *aph* gene confers kanamycin resistance and was amplified from the pKD4 plasmid (Table 2.2) with primers that have homology to *aph* and the gene of interest in the bacterial strain (Table 2.3). For the inactivation of *eefD*, the *aph* gene was amplified from pKD4 with primers P488 and P489. Deletion of the *eefABC* operon was achieved in a two-step process. Due to the size of the operon, *eefB* was deleted first, followed by the remaining *eefAC* region. The *eefB*::*aph* cassette was amplified using P394 and P395, *eefAC::aph* was amplified using P396 and P397. Amplimers were electrophoresed, purified and quantified as described in 2.6.

2.9.1.2 Introduction of the *aph* to the bacterial chromosome

The resistance cassette was incorporated into the bacterial chromosome using λ-red recombination. First pSLTS was isolated from EC23. Background strains were then transformed using pSLTS, a temperature inducible plasmid that encodes the λ-red recombinase, and recovered for 1 hour at 30˚C. Cells were then plated onto LB supplemented with 50 µg/mL ampicillin and recovered for 48 hours at 30˚C.

When constructing both Δ*eefB* and Δ*eefABC,* a single colony of *E. coli* ATCC 25922 (EC112) transformed with pSLTS was selected and used to inoculate 5 mL of LB broth supplemented with 50 µg/mL ampicillin and grown overnight 30˚C with aeration. The following morning, the strain was sub-cultured into LB broth supplemented with 50 µg/mL ampicillin and 2 mM arabinose and grown to OD_{600} 0.6. Cells were then washed in ice-cold 15% (v/v) glycerol and made electrocompetent as described in 2.3 before being transformed with the *aph* encoding amplimer. Following transformation, cells were recovered at 30˚C for three hours before being plated onto LB agar supplemented with kanamycin at a final concentration of 50 µg/mL. Candidate colonies were inoculated in 5 mL LB broth and grown overnight at 37˚C with aeration resulting in pSLTS curing. The following morning, a loopful of culture was spread over a prewarmed LB agar without antibiotics to isolate single colonies. Single colonies were then screened on LB, LB supplemented with 50 µg/mL ampicillin and LB supplemented with 50 µg/mL kanamycin. Colonies that could grow on LB and kanamycin but not ampicillin were taken forward for PCR and sequencing as described in 2.7 to confirm gene interruption.

Once gene interruption had been confirmed, the removal of the *aph* cassette was achieved using pCP20 which encodes an inducible recombinase. Cells containing the *aph* cassette were made electrocompetent as in 2.3 and transformed with pCP20. Cells were then recovered at 30˚C with aeration for 1 hour before being spread onto prewarmed LB agar supplemented with 50 µg/mL ampicillin and incubated at 30˚C for 24-48 hours. Candidate colonies were then cured of the pCP20 plasmid via growth at 43˚C in LB broth overnight with aeration. Following overnight growth, a loop of cultured was streaked onto LB agar and incubated at 37˚C overnight. Single colonies were

screened on LB, LB supplemented with ampicillin, and LB supplemented with kanamycin. Colonies that grew only on LB were then confirmed by PCR for both the loss of the *aph* cassette and pCP20 plasmid.

2.9.2 Bacteriophage transduction

2.9.2.1 Inactivation of *acrF* via P1 transduction

The inactivation of *acrF* in *E. coli* K-12 (EC07) and *E. coli* K-12 Δ*acrB* (EC29) was achieved through P1 phage transduction. A 5 mL overnight culture of *E. coli* K-12 BW25113 *acrF::aph* (EC113) (Baba *et al.*, 2006) was grown overnight in LB broth at 37˚C with aeration. The following day, 50 µL of overnight growth was used to inoculate 5 mL of LB broth supplemented with 50 µL 20% (w/v) glucose and 25 µL 1 M calcium chloride and incubated in a glass flask for 45 minutes at 37˚C with shaking. After 45 minutes of growth, 100 µL of *E. coli* K-12 MG1655 P1 *vir* stock (from Prof. Pete Lund) at a concentration of 10^9-10^{10} plaque forming units per mL was added to the culture which was then incubated for a further 3 hours at 37˚C until the bacterial cells had lysed. Following this, 2 drops of chloroform were added and the culture was incubated for a further 5 minutes. Supernatant was then transferred into a 50 mL Falcon and pelleted for 10 minutes at 2986 x *g.* Following centrifugation the chloroform layer was visible and the supernatant was carefully transferred into a 5 mL glass screw top vial, labelled as "P1 *vir acrF::aph*" and stored at 4˚C.

Following preparation of the EC113 *acrF::aph* P1 *vir* stock, 5 mL aliquots of LB broth were inoculated with *E. coli* K-12 MG1655 (EC07) and *E. coli* K-12 MG1655 Δ*acrB* (EC29) and grown overnight at 37˚C with aeration. The following day, 1.5 mL of each culture was pelleted for 2 minutes at 18,407 x *g* and resuspended in 0.75 mL of

P1 lysate salts solution (10 mM calcium chloride, 5 mM magnesium sulphate). For both *E. coli* K-12 MG1655 (EC07) and *E. coli* K-12 MG1655 Δ*acrB* (EC29), 100 µL of the bacteria/P1 salts mixture was mixed with 1 µL, 10 µL or 100 µL of EC113 P1 *vir* stock and incubated at room temperature for 30 minutes. A control of P1 *vir* stock without bacterial cells was also incubated for 30 minutes. Following the incubation step, 1 mL of LB broth supplemented with 200 µL of 1 M sodium citrate was added and these phage/bacteria mixtures were incubated at 37˚C for 1 hour with aeration. Cultures were then pelleted at 18,407 x *g* for 2 minutes. Pellets were resuspended in 50 µL of prewarmed LB broth and spread onto LB agar plates supplemented with 50 µg/mL kanamycin and 5 mM sodium citrate, the latter to help prevent P1 phage reabsorption. Bacteria and phage only controls were also spread onto LB agar supplemented with kanamycin and sodium citrate. Plates were incubated overnight at 37˚C. The number of colonies for each transduction was recorded and where colonies were present for more than one P1 *vir* stock concentration, colonies were isolated from those from the lowest P1 volume i.e. if there was growth for both for *E. coli* K-12 (EC07) mixed with 1 µL EC113 *acrF::aph* P1 *vir* stock and 10 µL EC113 *acrF::aph* P1 *vir* stock, colonies were taken from the 1 µL plate to minimise the presence of any P1 phage particles. Colonies were subsequently streaked to purity twice. The transduction of *acrF::aph* into the *E. coli* K-12 and *E. coli* K-12 Δ*acrB* backgrounds was confirmed by PCR and sequencing as described in 2.7.

2.10 Phenotypic characterisation assays

2.10.1 Ciprofloxacin accumulation

A single colony was inoculated into 5 mL of LB broth (supplemented with 50 µg/mL kanamycin or ampicillin if needed) and grown overnight at 37˚C with aeration. Using a 500 mL glass Duran bottle, 100 mL of LB broth was inoculated using 4% (v/v) overnight growth. Again, if the strain was carrying a plasmid 50 µg/mL kanamycin or ampicillin was added. The bacterial cultures were grown at 37˚C with aeration until an OD_{600} 0.6 was reached. The strain was then split into two 50 mL Falcon tubes and pelleted at 3900 x *g* for 10 minutes at 4˚C. The supernatant was removed and the pellet was washed twice with 10 mL pre-chilled 0.1 M potassium phosphate buffer (PPB) (3.08 mL 1 M dibasic potassium phosphate K_2HPO_4, 1.93 mL 1 M potassium phosphate monobasic KH_2PO_4, 45 mL sterile distilled water, pH adjust to 7). The pellet was then resuspended in 10 mL PPB and the OD_{600} was measured. The culture was then adjusted to OD_{600} 0.6. When using multiple strains, if one had an OD_{600} lower than 0.6, all cultures were diluted to the lowest OD_{600} in 10 mL.

Prior to the addition of ciprofloxacin, 100 µL of culture was removed for determining the CFU/mL. The remaining 9.9 mL of bacterial culture was incubated at 37˚C in a shaking water bath for 10 minutes. Following this incubation, 500 µL was removed and 94 µL of 1,000 µg/mL ciprofloxacin stock was added to the culture. Further 500 µL samples were removed 5, 10 and 20 minutes post the addition of ciprofloxacin. All samples taken were immediately incubated on ice.

Samples were then pelleted at 12470 x *g* for 1 minute at 4˚C. The supernatant was removed and pellets resuspended in 1 mL PPB. This process was done twice. Pellets were then resuspended in 1 mL 0.1 M glycine pH 3 and incubated at room

temperature for 2 hours in the dark. Following this incubation period, cells were pelleted at 12470 x *g* for 1 minute. Supernatant was then transferred to a new reaction tube and pelleted again at 12470 x *g* for 1 minute. In triplicate, 200 µL of each sample (0, 5, 10 and 20 minutes post-ciprofloxacin addition) were loaded to a black 96 well microtiter plate, as was 200 µL of 0.1 M glycine pH 3 buffer. The fluorescence of intracellular ciprofloxacin was measured using a FLUOstar Omega (BMG, UK) plate reader at excitation 280 nm and emission 440 nm. To quantify fluorescence, ciprofloxacin at known concentrations (0.0001 µg/mL to 1 µg/mL) was also added to the 96 well plate.

2.10.2 Quantification of biofilm production

LB broth was made without salt (10g tryptone (Oxoid; Cat. No. LP0042), 5g yeast extract (Oxoid; Cat. No. Cat. No. LP0021) and 1000 mL H_2O) and used to culture strains for this experiment. Overnights were prepared from single colonies and grown overnight at 37˚C with aeration. The following morning cultures were adjusted to OD_{600} 0.1 in LB without salt and 200 µL of each strain was aliquoted into a round bottom 96-well microtiter plate in triplicate (technical repeats). Plates were incubated at 30˚C for 48 hours with gentle agitation.

After 48 hours, microtiter wells were washed with water to removed unattached cells before 200 µL of 0.1% (w/v) crystal violet was added. The plate was incubated at room temperature for 15 minutes, crystal violet was then removed and wells were washed again with water to remove excess crystal violet. Crystal violet bound to the biofilm was solubilised using 200 µL 70% (v/v) ethanol, the optical density was measured at a wavelength of 600 nm and quantified. *Salmonella enterica* subsp.

enterica serovar Typhimurium 14028s was used as a positive control for biofilm production.

2.10.3 Measurement of antimicrobial susceptibility by agar dilution method

The antimicrobial susceptibility of dyes and antimicrobials was determined using the agar dilution method as described by the Clinical and Laboratory Standards Institute (CLSI, 2020). Antimicrobial and dye stocks were prepared on the day and used to supplement Iso-sensitest agar (Oxoid, UK; Cat. No. CM0471) (Table 2.5). The required volume of compound for each concentration tested was mixed with 20 mL agar which was dispensed by a peristaltic pump (Jencons Scientific Ltd., UK). Supplemented agar was then poured into a 100 mm square petri dish (Thermo Fisher) and allowed to set, plates were then dried at 60˚C for 15 minutes. The optical density of overnight cultures was measured and strains were adjusted to density of the culture giving the lowest OD_{600}. All strains were then diluted 1:100 and 180 µL per strain was aliquoted into a 96 well round bottom microtiter plate. *E. coli* ATCC 25922 (EC112) was used as a control strain, as this strain has known MIC values and confirms compound efficacy (EUCAST, 2021). Once dry, agar plates were inoculated with 1 µL of each strain using a multipoint inoculator (Mast Uridot, Mast Group Ltd.). Inoculated plates were left to dry before being incubated at 37˚C overnight. The minimum inhibitory concentration (MIC) of compounds was defined as the first concentration in which the bacterial strain did not grow.

Table 2.5 Compounds used in MIC

Compound	Supplier	Catalogue number	Solvent	Concentration range (µg/mL)	
Aztreonam	Fisher	455410010	DMSO	0.008	4
Benzalkonium Chloride	Alfa Aesar	11384048	SDW	1	128
Carbenicillin disodium salt	Fisher	15414429	SDW	1	1024
Cefotaxime sodium salt	Fisher	10084487	SDW	0.015	256
Chloramphenicol	Sigma	C0378	70% Methanol	0.008	128
Ciprofloxacin	Acros	13531640	SDW + acetic acid	0.008	128
Clindamycin hydrochloride	Generon	A10227	DMSO	0.125	256
Crystal violet	Fisher	10646972	SDW	0.25	128
Erythromycin	Fisher	10338080	70% Ethanol	0.5	1024
Ethidium bromide	Fisher	10042120	SDW	1	1024
Fusidic acid sodium salt	Sigma	F0881	SDW	1	1024
Gentamicin sulfate	Fisher	10224873	SDW	0.008	32
Meropenem trihydrate	TCI Chemicals	M2279	SDW	0.008	32
Methylene Blue	Fisher	10307033	SDW	1	1024
Moxifloxacin hydrochloride	Sigma	PHR1542	60% DMSO	0.008	32
Nalidixic acid sodium salt	Fisher	1037362	SDW	1	1024
Novobiocin sodium salt	Fisher	15403619	SDW	1	512
Rhodamine 6G	Sigma	252433	SDW	1	1024
Rifampicin	Fisher	10533325	SDW	0.25	32
Spectinomycin	Sigma	S9007	SDW	0.5	16
Tetracycline hydrochloride	Fisher	10460264	SDW	0.015	1024
Ticarcillin disodium salt	Sigma	T5639	SDW	0.125	1024

2.10.4 Measurement of curli and cellulose production using Congo red

Overnights were grown at 37˚C with aeration in LB broth. LB agar without salt (5 g tryptone (Oxoid; Cat. No. LP0042), 2.5g yeast extract (Oxoid; Cat. No. LP0021), 7.5g agarose (Oxoid; Cat. No. LP0011) in 500 mL of distilled water) and a Congo red (Alfa Aesar; Cat. No. B24310) stock at a concentration of 10,000µg/mL (0.1g in 10mL SDW, and then filter sterilised) were both prepared. After autoclaving the agar and allowing it to cool to ~50˚C, Congo red was added at a final concentration of 40 µg/mL. Plates were then dried at 60˚C before inoculation.

The OD_{600} of overnight cultures was measured and strains were then adjusted a final OD_{600} of 0.1. If strains were carrying a plasmid, antibiotics were added in the final dilution to aide selection. For each strain 5 µL was spotted onto the plate in triplicate, allowed to dry and plates then incubated for 48 hours at 30˚C. Rough red colonies indicated the production of both curli and cellulose, pink indicated strains produce cellulose not curli, and smooth white indicated production of neither. Colonies were imaged with a SMZ1000 zoom stereomicroscope (Nikon, Japan).

2.10.5 Ethidium bromide efflux

Overnight cultures were grown at 37˚C with aeration in LB broth (supplemented with ampicillin or kanamycin if carrying plasmids). 10 mL LB broth was then inoculated with 4% (v/v) of overnight culture and incubated at 37˚C with aeration until to an OD_{600} of 0.4 was reached. Cells were then pelleted at 2986 x *g* at 23˚C for 10 minutes and resuspended in 10 mL 20 mM potassium phosphate buffer (PPB) supplemented with 1 mM $MgCl_2$. To make this buffer, first a 0.1 M PPB (pH 7) stock was made from 3.08 mL 1 M K_2HPO_4, 1.93 mL 1 mL KH_2PO_4, and 45 mL sterile distilled water. To produce

20 mM PPB + 1 mM $MgCl_2$, 50 mL 0.1 M PPB stock was mixed with 200 mL sterile distilled water and 250 µL of 1 M $MgCl_2$

Strains were adjusted to an OD_{600} of 0.2 in 10 mL 20 mM PPB with 1 mM $MgCl_2$. Next 100 µL of a 10 mM carbonyl cyanide m-chlorophenyl hydrazone (CCCP) stock (0.01 g in 4 mL SDW;1 mL DMSO) and 500 µL of a 1000 µg/mL ethidium bromide stock (0.01 g in 10 mL SDW) were added to each culture and strains were incubated at 23˚C with aeration for 1 hour. Following the hour incubation, cultures were once again pelleted for 10 minutes at 23˚C and 2986 x *g*. Pellets were then resuspended in 9 mL of 20 mM PPB + 1 mM $MgCl_2$. In triplicate 190 µL of each strain was aliquoted into a black 96-well microtiter plate. The plate was then loaded into the Fluostar Omega (BMG, UK) reader and fluorescence intensity was measured for 100 minutes at excitation 544 nm and emission 590 nm, gain was set to 1460. After the first plate measurement, cells were reenergised with 10 µL of 20% glucose (w/v) (final concentration 1%) using the FLUOstar injector. The number of samples and technical replicates loaded to the plate was kept to a minimum to ensure cycle time was as quick as possible. For each experiment, three technical replicates were measured and the mean time taken for fluorescence to decrease by 10%, 25% and 50% was calculated for each strain.

2.10.6 1,2'-dinaphthylamine efflux

During this study a method to measure the rate of 1,2'-dinaphthylamine (1,2'-DNA) efflux was established. This method was adapted from (Philippe *et al.*, 2015). Unlike ethidium bromide, 1,2'-DNA does not associate with DNA and instead accumulates within the periplasm.

Single colonies were used to inoculate 5 mL LB broth (supplemented with 50 µg/mL ampicillin or kanamycin if needed), and cultures were grown overnight at 37˚C with aeration. Overnights were sub-cultured into fresh LB broth (with kanamycin or ampicillin if strains were carrying a plasmid) and incubated at 37˚C until an OD_{600} of 0.4 was reached. Cultures were then harvested by centrifugation at 2986 x *g* for 10 minutes and resuspended in 10 mL 20 mM PPB with 1 mM $MgCl_2$ and adjusted so all cultures were of OD_{600} 0.25 in a final volume of 10 mL and stored on ice. A 100,000 µM stock of 1,2'-DNA (TCI Chemicals, Japan; Cat. No. D2988) was prepared by dissolving 0.269 g in 10 mL DMSO. This was then diluted to 1000 µM in 80% DMSO. A 5 mM CCCP stock was prepared by dissolving 0.01 g in 8 mL DMSO and 2 mL sterile distilled water. Following OD_{600} adjustment to 0.25, 320 µL of 1000 µM 1,2'-DNA and 10 µL were added and cultures were incubated in the dark at 37˚C overnight with aeration.

The following morning, cells were pelleted by centrifugation at 2986 x *g* for 10 minutes and then washed in 10 mL 20 mM PPB with 1 mM $MgCl_2$ twice. After washing, pellets were resuspended in 10 mL 20 mM PPB with 1 mM $MgCl_2$ and 190 µL of each culture was aliquoted into a black microtiter plate and fluorescence intensity was read at 37˚C using a FLUOstar Omega (BMG). Gain was 1830, and excitation and emission were set at 370-10 nm and 420-10 nm, respectively. Fluorescence was read using "top reading – plate mode", with cycles measured using orbital averaging and a scanning diameter of 2 mm (10 flashes per well). Settling time was set to 0.1 s and cycle time was minimised by having only one technical replicate per culture on the plate during the assay. Efflux of 1,2'-DNA was reinitialised by 10 µL injection of 1 M glucose during read cycle 2. The injection step was performed using the built-in FLUOstar injector.

Fluorescence intensity was measured over 10 minutes. A further 2 plates were loaded with OD_{600} adjusted cultures that had been on ice and the fluorescence intensity was measured over a 10 minute period following cell re-energisation to provide a second and third technical replicate.

As starting fluorescence values varied between biological replicates, measurements were normalised to start at 1. Mean fluorescence values were determined for the three technical and biological replicates, and the time taken for fluorescence intensity to drop by 10% and 20%

2.11 Statistics

All experiments were performed a minimum of three times. The number of technical/biological replicates is displayed in the figure legend along with the statistical test used. Statistical tests were performed using GraphPad Prism (v 9). A p value of ≤ 0.05 was considered to be significant in all tests.

2.12 Bioinformatics

2.12.1 Chapters 1, 2 and 3

2.12.1.1 Pangenome analysis

Prior to this project, genome assemblies of 20 *E. coli* sequence types were downloaded from Enterobase (Zhou *et al.*, 2020) and pangenomes were generated using Roary (Page *et al.*, 2015) by Christopher Connor (CC). The genes of 15 RND genes from *E. coli* K-12 MG1655 (NC_000913.3) (Appendix 8.1) were then aligned using BLASTN to the pan_genome_reference.fa file generated by Roary. Pangenome

alleles were defined as hits with ≥90% sequence coverage, ≥90% sequence identity and ±0.5% K-12 sequence length.

2.12.1.2 Identifying SNPs across RND proteins

The genome assemblies used to generate 20 *E. coli* pangenomes were aligned to the peptide sequences of 15 RND genes (Appendix 8.1) using TBLASTN. TBLASTN searches a translated nucleotide sequence in all six frames using a known protein query. The genetic code was specified as bacteria and archaea (11) and the "seg" flag was set to "no". This resulted in 15 BLAST results files for each of the 20 STs, results were then filtered by length and by a string. The string contained amino acids 2 to 6 of the protein sequence. Filtering by amino acid sequence allowed the separation of proteins that matched in length e.g. MdtF and AcrD. The first amino acid was not included in the string to allow for non-methionine start codons.

Once filtered, K-12 protein sequences were concatenated to each of the 300 files and these were aligned using Muscle (v 3.8.1551) (Edgar, 2004). Due to the number and length of AcrB sequences present the ST11 genomes, the server was unable to perform more than two iterations of Muscle, therefore a maximum number of 2 iterations was used for all alignments. Asparagine residues in alignment files were replaced with a "Z" prior to running SNP-sites as "N" was ignored by the software. SNP-sites (v 2.5.1) (Page *et al.*, 2016) was then run and the resulting VCF converted to an Excel file to allow analysis. The total number of "SNPs" (here amino acid polymorphisms) at each residue was determined for all 300 proteins. Heatmaps to visualise the final data were generated using heatmaps.2 of the gplots package (v 3.1.1) on R (v 4.0.4).

2.12.1.3 Conservation of AcrF insertion across ST11

As no *acrF* gene alleles were identified across the ST11 pangenome, the region downstream of *acrE* was downloaded from 13 annotated *E. coli* ST11 genomes available on NCBI (NC_000913.3, NC_011353.1, NC_013008.1, NZ_CP008805.1, NZ_CP010304.1, NC_002655.2, NZ_CP008957.1, NZ_CP012802.1, NZ_CP015846.1, NZ_CP015842.1, NZ_CP015843.2, NZ_CP016625.1, NC_002695.2). These regions were then aligned using Muscle and alignments were visualised using Unipro UGENE (v 37.0) (Okonechnikov *et al.*, 2012).

To determine the prevalence of the 45 nucleotide across ST11, 2000 assemblies were downloaded from Enterobase, duplicates were removed as in 2.12.2 and the *acrF* gene sequence containing the insertion from *E. coli* TW14359 (NC_013008.1) was aligned to ST11 assemblies using BLASTN.

2.12.2 Chapter 4

2.12.2.1 Abricate analysis

The presence of *eefB* across *E. coli* (Taxid 562) genomes available on NCBI was determined using the *eefB* nucleotide sequence from *E. coli* SMS-3-5 (CP000970.1; gene EcSMS35_183). Once it was apparent *eefB* was prevalent across *E. coli,* Abricate was used. Abricate (Seemann, 2017) is programme that enables the screening of contigs for antimicrobial resistance and virulence genes as it utilises databases such CARD (Alcock *et al.*, 2019). Abricate also supports the use of custom databases, allowing contigs and assemblies to be quickly screened for any genes.

Non-legacy genome assemblies of 38 *E. coli* STs were downloaded from Enterobase (Zhou *et al.*, 2020) using a script written by CC (available at https://github.com/C-Connor/EnterobaseGenomeAssemblyDownload). Duplicates were then removed using a custom script also written by CC (available at https://github.com/C-Connor/MashDistDeReplication). A custom database was built using the *eefRABCD* genes from *E. coli* ATCC 25922 (accession CP009072.1) (Appendix 8.2) and makeblastdb from the BLAST package (v 2.10.0) (Camacho *et al.*, 2009). Abricate (v 0.9.8) (Seemann, 2017) was then run to determine gene presence and conservation across the genome assemblies using the *eefRABCD* database. Those found to have an *eef* gene or genes split over multiple contigs were removed from the analysis.

2.12.2.2 Phylogenetic tree construction

For each ST, five genome assemblies were randomly selected for generating the phylogenetic tree. To do this the “shuf” command was used. Assemblies were then annotated using Prokka (v 1.14.6) (Seemann, 2014). To speed up this process the autoprokka.py script written by Steven Dunn was used with the following flags “genus --Escherichia”, “species --coli” and “--compliant” (available at https://github.com/stevenjdunn/autoprokka).

All available *Escherichia fergusonii* assemblies were downloaded from Enterobase (*n* = 27) and duplicates were removed (final *n* = 24). As with *E. coli,* five random assemblies were selected using the “shuf” command. *E. fergusonii* assemblies were annotated using autoprokka.py this time using the “genus --Escherichia”, “species --fergusonii” and “--compliant” flags.

A simple phylogenetic tree was generated using ParSNP (v 1.2) (Treangen *et al.*, 2014). As this suggested one of the *E. fergusonii* assemblies was in fact *E. coli,* the species of assemblies used in the analysis were confirmed using Kraken2 (v 2.0.8-beta) (Wood *et al.*, 2019). ST group and phylogroup were also checked using MLST (v 2.18.0) (available at https://github.com/tseemann/mlst) and ClermonTyping.sh (v 1.4.1) (Beghain *et al.*, 2018) respectively.

Following confirmation of species, ST and phylogroup, the GFF files generated by Prokka were used to run Roary (v 3.13.0) (Page *et al.*, 2015) with the following command "-e -i 80 --verbose --mafft -p 4". RAxML (v 8.2.12) (Stamatakis, 2014) was used to generate the final phylogenetic tree using the following flags "raxmlHPC-PTHREADS-AVX -T 4 -f a -p 12345 -s core_gene_alignment.aln -x 12345 -# 100 -m GTRGAMMA -n raxml.out", the core_gene_alignment.aln file having been generated by Roary. The generated tree was then visualised and edited using iTOL (Letunic and Bork, 2019).

2.12.2.3 Determining the genomic context of *eefRABCD*

The genomes of seven reference sequences were downloaded from the NCBI RefSeq database. These belonged to four sequence types that were identified to have the *eef* operon and three that did not. NC_004431.1 (ST73), NC_007946.1 (ST95), NZ_HG941718.1 (ST131), NC_002695.2 (ST11) all possessed the *eef* operon, while NC_000913.3 (ST10), NC_011751.1 (ST69), and NZ_CP035350.1 (ST617) did not. The location of the *eef* operon was identified in ST73, ST95, ST131 and ST11 genomes using BLASTN alignment. The region was then downloaded plus 10,000 bases down- and upstream of the region in GenBank format. Once apparent that *fabI*

and *sapF* flanked the operon, these genes were located in the genomes that lacked the *eef* operon and 10,000 bases down- and upstream of *fabI* and *sapF* were downloaded from each of the ST10, ST69 and ST617 genomes. Easyfig (v 2.2.2) was used to visualise the homology between *eef*-encoding and *eef*-lacking strains. Gene annotation in figures was sourced from feature labels within GenBank files.

2.12.2.4 Identifying genes that co-occur with *eefRABCD*

To identify genes that statistically co-occurred with the *eef* operon, Coinfinder (v 1.1) (Whelan *et al.*, 2020) was used. The random sample of genome assemblies used to produce the phylogenetic tree were once again used in this analysis. The gene_presence_absence.csv generated by Roary prior to running RAxML was also used to run Coinfinder. The required Newick phylogenetic tree was generated with Mashtree (v 1.0.4) (Katz *et al.*, 2019) using the fasta files of the genomes used to generate to the RAxML phylogenetic tree. The resulting GEXF files were visualised using Gephi (v 0.9.1) (Mathieu *et al.*, 2009) and each of the *eef* genes and their neighbouring nodes (linked genes) were isolated. Node colours and shapes were ranked by the *D*-value i.e as the *D*-value decreased as did the node diameter, furthermore node colour was also linked to the *D*-value. Nodes with a *D*-value closer to 0 were green while those closer to -1 were pink. Map layout was defined using the Fructerman Reignhold model settled around the *eef* gene of interest.

2.12.2.5 Determining the presence of *eef* across further species

As in 2.12.2.1, *Shigella* assemblies were downloaded from the Enterobase database. A total of 500 assemblies were downloaded per *Shigella* species; *Shigella boydii,*

Shigella dysenteriae, Shigella flexneri, and *Shigella sonnei.* Duplicates were identified and removed from the analysis as in 2.12.2.1. Due to the genetic similarity between *E. coli* and *Shigella,* the same *eefRABCD* custom database was used to run Abricate. Again assemblies with single *eef* genes split over multiple contigs were removed from the analysis.

To determine the genomic context of *eefRABCD* in *S. dysenteriae,* 15 genome assemblies were selected randomly. These files were annotated using Prokka "--genus Shigella --species dysenteriae --compliant". The reference sequence *Shigella dysenteriae* ATCC 12039 (CP026774.1) was also annotated using Prokka. The presence of insertion sequences across the *eef* operon was suggested by Prokka annotation. As a result ISEscan (v 1.7.2) (Xie and Tang, 2017) was used to further identify any insertion sequences.

The presence of *eefRABCD* across further species was initially determined using NCBI BLAST. *Photorhabdus* (Taxid 29487)*, Salmonella* (Taxid 590)*, Yersinia* (Taxid 629), *Serratia* (Taxid 613), *Pseudomonas aeruginosa* (Taxid 287) and *Acinetobacter baumannii* (Taxid 470) were searched using the Taxid filter with *eefR, eefA, eefB, eefC* and *eefD* as the query from *E. coli* ATCC 25922.

To determine the genomic context of the *eef* operon in *Escherichia marmotae* (CP056165.1)*, Escherichia albertii* (CP070290.2)*, Klebsiella pneumoniae* (NC_016845.1)*,* and *Klebsiella aerogenes* (NZ_CP041925.1) in relation to *E. coli* (CP009072.1)*,* a single RefSeq genome was downloaded per species from NCBI in the GenBank format. Comparisons were visualised using EasyFig.

Chapter 3 Conservation of RND genes and proteins across 20 *Escherichia coli* sequence types

3.1 Background

Bacterial efflux pumps are known to perform a wide range physiological functions, including oxidative and nitrosative stress protection, mammalian bile salts export, cell to cell signalling, and the export of antimicrobial compounds such as antibiotics (Henderson *et al.*, 2021).

The ever increasing availability of whole genome sequences for a wide range of bacterial organisms means it is now possible to determine the conservation of specific genes across thousands of genomes. With the development of bioinformatic tools and pipelines, one can also determine whether particular genes are present in the core or accessory genome, indicating their cellular importance.

While amino acid substitutions in RND efflux pumps have been reported, they are frequently discovered in single isolates following exposure to antimicrobials which results in a selection pressure (Blair *et al.*, 2015a, Elkins *et al.*, 2010). Moreover, much of the work on RND structure is done using well-characterised laboratory strains. To our knowledge the conservation of RND protein sequences has only be studied between species, using single representatives of a species and therefore the true conservation of RND structures is not known (Tseng *et al.*, 1999, Zgurskaya *et al.*, 2009).

3.2 Aims

- Determine the conservation of RND efflux pump genes across *E. coli* pangenomes
- Assess the level of protein conservation at the individual assembly level for the 20 *E. coli* ST

3.3 Hypotheses

- The gene sequences of RND pumps in *E. coli* K-12 will be conserved across *E. coli*
- Pumps with the most important biological functions, such as AcrAB-TolC, will have the highest level of conservation

3.4 Identification of RND genes across *E. coli* pangenomes

The initial aim of this thesis was to understand how conserved RND genes are across *E. coli.* To do this the number of alleles per each RND-associated gene was determined across 20 *E. coli* ST pangenomes.

Prior to this project, 19,601 genome assemblies were downloaded and pangenomes for 20 *E. coli* STs belonging to 7 phylogroups were created by CC using Roary, a prokaryote pangenome construction tool. The STs selected spanned ExPEC, EPEC and EHEC, and lineages associated with commensal and more variable niches.

One of the outputs of Roary is a fasta file that contains a single representative nucleotide sequence for each cluster. Here a cluster refers to a group of nucleotide sequences that share greater than 95% identity at the protein level. Clusters are determined by an all-against-all BLASTP (all sequences within the analysis compared to one another) alignment set with a minimum sequence identity percentage; this can

either be specified or set at 95% if the programme is run with default conditions. A single nucleotide sequence is then selected by the programme to represent this protein cluster. This information is then stored in a pangenome reference file (pan_genome_reference.fa).

Using the nucleotide sequences of genes encoding RND pumps and their associated PAP and OMP from *E. coli* K-12 (accession NC_000913.3), RND-associated gene alleles were identified within each pangenome reference file (pan_genome_reference.fa) using BLASTN. An allele was defined as such if it had a nucleotide identity percentage ≥90% and a sequence length within ±0.5% of the *E. coli* K-12 sequence. In the first instance a percentage identity value of ≥95% was used but this resulted in *mdtC* alleles across several lineages being excluded from the analysis as nucleotide identity can be as low as 93%. Therefore, a percentage identity cut off of greater than 90% was used throughout. A sequence length of ±0.5% of the known length of *E. coli* K-12 sequences was chosen to allow for subtle variations in nucleotide sequences i.e. insertions and deletions.

3.5 The number of RND gene alleles varies between *E. coli* pangenomes

Across the 20 STs, RND genes were generally highly conserved as a low number of gene alleles were present across the 20 pangenomes, regardless of phylogroup the lineage belonged to. The most conserved RND systems were the AcrAB and MdtEF and their associated OMP, TolC (Table 3.1). Due to its clinically important role in drug resistance and proven role in biofilm formation and virulence, genes encoding AcrAB-TolC were expected to be well conserved. Both the PAP, *acrA*, and RND pump, *acrB*, had an average 1.2 alleles per ST showing high levels of conservation. Across the 20

STs only two lineages had more than one allele for *tolC.* TolC is the OMP of five of the six RND systems in *E. coli,* as a result any significant variations in the nucleotide sequence e.g. insertions and deletions would impact on the efflux capabilities of all RND efflux systems except for CusABC. It is therefore unsurprising the mean number of alleles was 1.1. It has been shown that the AcrAB-TolC complex has a fourth component, AcrZ, which is an inner membrane protein that modulates the substrate range of the pump (Hobbs *et al.*, 2012). Though this 150 nucleotide gene was found to be non-essential in K-12 (Goodall *et al.*, 2018), across the 20 ST only a single allele was found per pangenome suggesting it is under very strong positive selective pressure.

MdtEF is not related to antibiotic resistance but it's importance in other facets of *E. coli* biology have been recently elucidated. MdtEF has been shown to increase pH tolerance and provide bacterial cells with protection against nitrosative stress caused during anaerobic respiration (Schaffner *et al.*, 2021, Zhang *et al.*, 2011). Across the 20 *E. coli* STs, both the PAP and RND pump were found to be highly conserved, with mean number of alleles being 1.1 and 1.3 for *mdtE* and *mdtF,* respectively.

The highest mean number of alleles for a single gene was 1.6 (*acrF* and *mdtC*), however, the most variable tripartite efflux systems across the 20 STs were *cusABC* and *mdtABC* which both had a mean of 1.45 alleles. CusABC is the only heavy metal RND efflux pump in *E. coli* and increases bacterial tolerance to copper, a metal that is a component of the human immune system (Ladomersky and Petris, 2015). MdtABC is the only RND pump in *E. coli* that is a heterotrimer, and has been shown to export fosfomycin and quinolones, as well as bile salts that are a component of the human digestive system (Nagakubo *et al.*, 2002, Nishino and Yamaguchi, 2001).

No alleles were identified for five genes in specific STs, these were *acrD* (ST3), *acrE* (ST167), *acrF* (ST11, ST38) and *mdtF* (ST372). In each case alignment of the *E. coli* K-12 nucleotide sequences to the pangenome reference file in question identified hits with >90% nucleotide identity. However, none were within the ±0.5% cut-off specified for gene length. In several cases, two hits were present which when combined were extremely close to the expected length of the gene in question suggesting that either a) the position of the gene on the assembly fragment had led to it being split over multiple contigs or b) the presence of an insertion or deletion which has affected the coding sequence (CDS) and subsequent annotation by PROKKA. To confirm the presence or absence of the genes from ST3, ST167, ST11, ST38 and ST372, the assemblies used to create the pangenome were aligned with the K-12 reference nucleotide sequences. The absence of identified alleles for both AcrF in ST38 and MdtF in ST372 were found to be artefacts as TBLASTN analysis done following analysis with Roary (explained in next section) identified the AcrF protein in 88.5% of the in 617 inputted ST38 assemblies. While 98.1% of the 54 ST372 assemblies have the MdtF protein. AcrD and AcrE were both found to be present in the assemblies of ST3 and ST167 respectively, but to a lesser extent. Only 32.5% and 18.3% of the 40 and 115 assemblies belonging to ST3 and ST167 had the protein of interest present. However, when looking at the prevalence of AcrF across ST11, TBLASTN identified the protein in only 0.54% of the 5137 assemblies used to generate the ST11 pangenome suggesting the gene was truly absent. Due to the virulence of ST11 and the described impact *S.* Typhimurium AcrF homolog on virulence, the absence of *acrF* from its pangenome was further studied in Chapter 4.

The data in Table 3.1 also shows that the variation of efflux systems differs across the STs. ST10 and ST131 had the greatest number of alleles or the most variable efflux genes, while ST3 and ST372 had the least variability. Both ST10 and ST131 were shown to have the largest pangenomes, therefore as ST10 and ST131 have the largest pangenomes in the data it is unsurprising that they have the most RND alleles (Connor, 2021).

Table 3.1 Pangenome alleles of RND-associated genes across 20 *E. coli* sequence types

			Resistance-Nodulation-Division gene														
Phylogroup	*n*	Lineage	*acrA*	*acrB*	*acrD*	*acrE*	*acrF*	*acrZ*	*cusA*	*cusB*	*cusC*	*mdtA*	*mtdtB*	*mdtC*	*mdtE*	*mdtF*	*tolC*
A	2370	ST10	1	1	2	2	5	1	3	3	4	1	2	5	3	2	2
A	115	ST167	1	1	1	0	1	1	1	1	1	1	2	2	1	1	1
B1	40	ST3	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
B1	1884	ST17	1	1	1	1	2	1	2	3	2	2	2	2	1	2	1
B1	2441	ST21	1	1	2	2	1	1	1	1	2	1	2	1	1	1	1
B2	283	ST12	1	2	1	2	4	1	1	1	2	1	1	1	1	1	1
B2	62	ST14	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
B2	46	ST28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
B2	873	ST73	2	2	2	1	1	1	1	1	2	1	3	2	1	1	1
B2	758	ST95	2	1	1	1	1	1	1	2	1	1	1	1	1	2	1
B2	232	ST127	2	1	2	1	1	1	2	2	1	2	1	1	1	1	1
B2	3186	ST131	1	2	2	2	2	1	4	3	2	5	3	4	1	2	1
B2	91	ST141	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
B2	65	ST144	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
B2	54	ST372	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
D	617	ST38	1	1	2	1	0	1	1	1	1	1	1	1	1	2	1
D	696	ST69	1	1	2	2	2	1	1	1	1	1	1	1	1	2	1
E	5137	ST11	2	2	1	1	0	1	2	2	3	2	1	2	1	2	1
F	269	ST117	1	1	1	1	2	1	1	1	1	1	1	2	1	1	2
F	382	ST648	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Total	19601	Mean number of alleles	1.2	1.2	1.3	1.2	1.6	1.0	1.4	1.5	1.5	1.4	1.4	1.6	1.1	1.3	1.1

An allele is defined here as a sequence with ≥ 90% sequence identity and within ± 0.5% of the expected nucleotide length in comparison to the *E. coli* K-12 reference sequence. Cells are highlighted red to indicate the number of gene alleles present.

3.6 Diversity in protein sequences at the individual ST and phylogroup level

To further understand how conserved RND systems are across *E. coli,* 15 *E. coli* K-12 RND and RND associated sequences were aligned with the assemblies used to construct the 20 pangenomes, this time using TBLASTN. Here the protein sequences were aligned with the translated nucleotide assemblies to assess protein conservation. Protein sequences were used instead of nucleotides as this removed any non-synonymous polymorphisms and only amino acid sequences changed by SNPs would be identified. Furthermore, as Roary had incorrectly suggested the absence of genes from multiple STs, it was decided that using the assemblies would give a higher accuracy when looking at the variation of RND pumps across in *E. coli.*

As mentioned, *E. coli* K-12 was used as the reference. While this strain has a considerably smaller genome compared to pathogenic *E. coli* lineages, it is a well characterised laboratory strain and was therefore used as the reference due to its frequent use in phenotypic and structural characterisation studies of efflux proteins (Wang *et al.*, 2017, Du *et al.*, 2014, Hayashi *et al.*, 2001). Furthermore, recently solved structures belong to either *E. coli* K-12 or derivatives of K-12 (Du *et al.*, 2014, Chen *et al.*, 2021). It is therefore useful to know whether the solved structures truly represent those of RND pumps found across *E. coli*.

3.6.1 AcrAB-TolC

Data showing the extent of conservation at each amino acid position in each of the RND and RND-associated proteins are presented as heatmaps.

AcrA, AcrB and TolC form the main RND system in *E. coli* and each protein was found to be highly similar to the sequence seen in *E. coli* K-12 as there was very little

variation across all 20 STs studied suggesting strong positive selection. However, each protein had a dominant polymorphism present at a single position. In AcrA this position was 104, here a polymorphism was present and highly conserved in phylogroups B2 (ST12, ST14, ST28, ST73, ST95, ST127, ST131, ST141, ST144, ST372), D (ST38, ST69) and F (ST648 and ST117) (Figure 3.1). In these ST groups the T104A substitution was present in >99% of assemblies.

In AcrB a dominant polymorphism was seen at amino acid 596 where histidine was changed to asparagine (Figure 3.2). However, in AcrB there isn't such a clear divide between the phylogroups. The K-12 amino acid of histidine was conserved at position 596 in phylogroup A (ST10, ST167), phylogroup B1 (ST3, ST17, ST21), phylogroup B2 (ST12, ST95), the only phylogroup E in the study (ST11), and phylogroup F (ST648 and ST117). While in the 8 of the 10 B2 lineages (ST14, ST28, ST73, ST127, ST131, ST141, ST144, ST372) and both phylogroup D STs (ST38 and ST69) the alternate amino acid of asparagine was present and conserved in 100% of assemblies.

In TolC, a polymorphism was seen at position 233 where instead of alanine found in the K-12 protein, strains possess a threonine (Figure 3.3). Variation from the K-12 sequence was seen in 9 of the 10 B2 STs (ST12, ST14, ST28, ST73, ST95, ST127, ST131, ST141, ST372), one phylogroup D (ST69) and one F (ST117).

Interestingly AcrA T104A, AcrB H596N and TolC A233T were found to be present in phylogroups B2, D and F but not A, B1 and E. The AcrAB-TolC amino acid variations will be discussed further in Chapter 5 including their position in the protein.

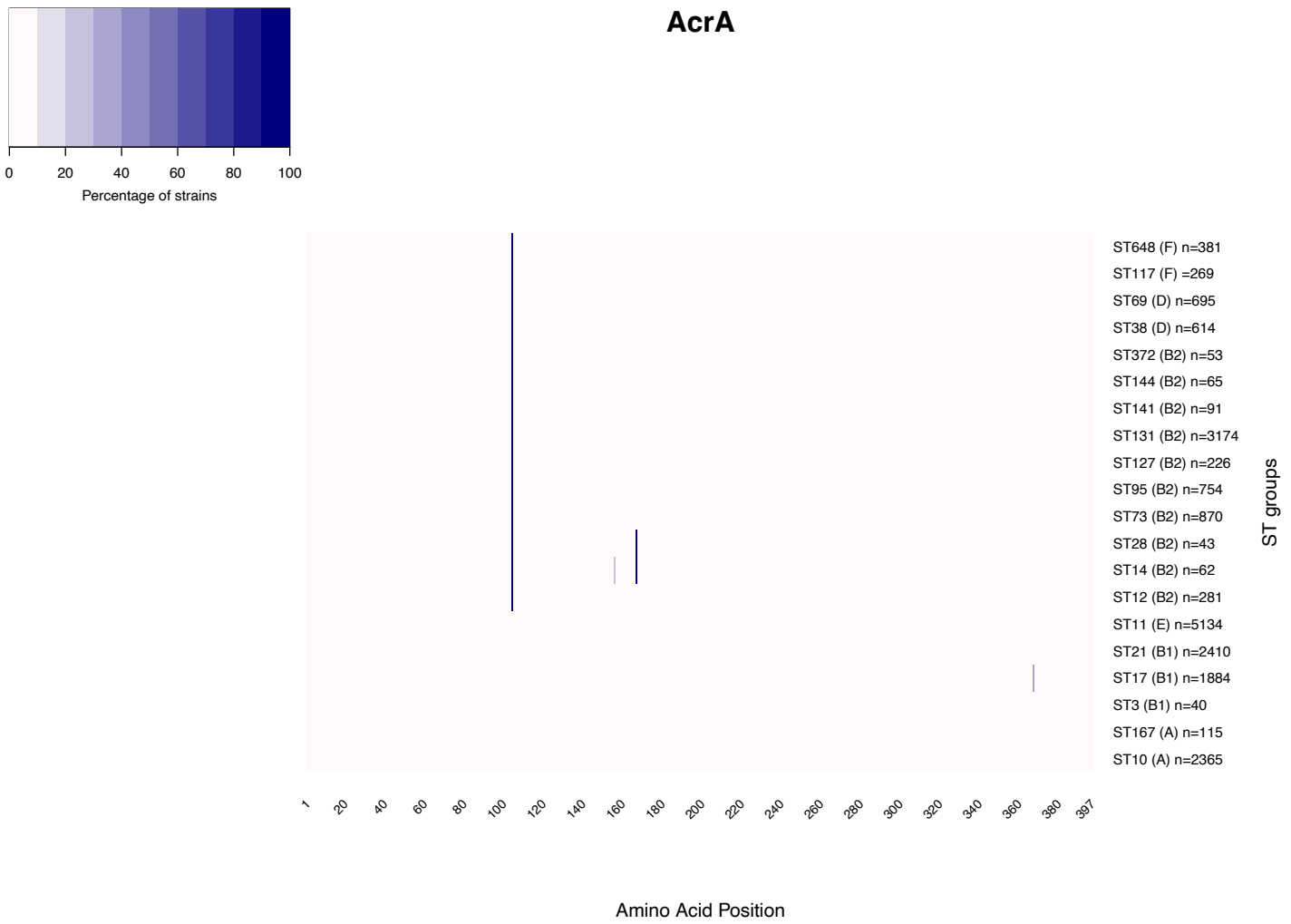


Figure 3.1 Conservation of the AcrA protein across 20 *E. coli* STs

The conservation of AcrA proteins across the 20 STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to AcrA (NP_414996.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

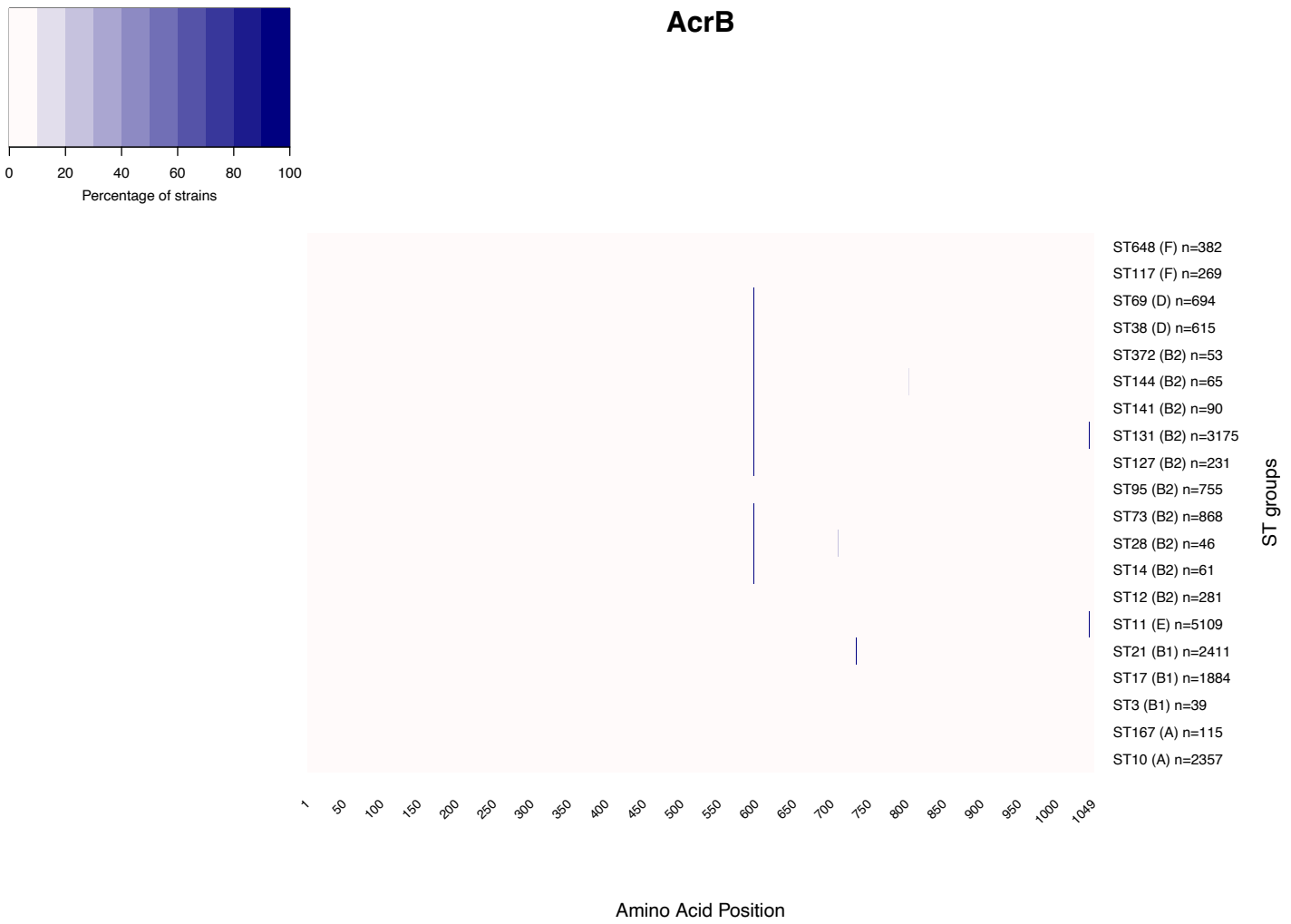


Figure 3.2 Conservation of the AcrB protein across 20 *E. coli* STs

The conservation of AcrB proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to AcrB (NP_414995.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

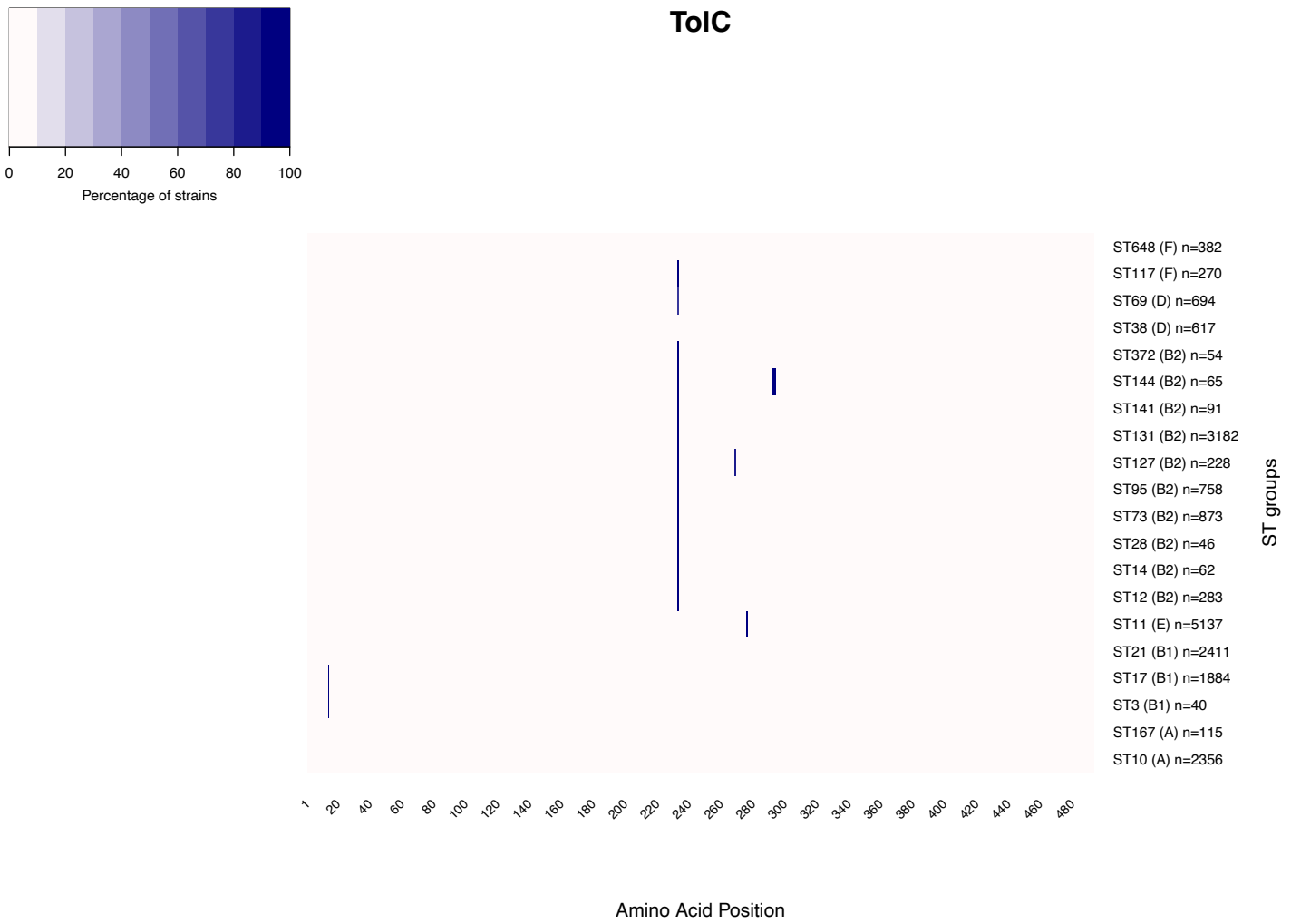


Figure 3.3 Conservation of the TolC protein across 20 *E. coli* STs

The conservation of TolC proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to TolC (NP_417507.2) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

3.6.2 AcrZ

Across the pangenome *acrZ* was the only gene in which a single allele was found per ST. Comparison of the protein sequences from the 20 STs to that of K-12 identified no amino acid changes across the individual 19,562 protein sequences identified using TBLASTN (Figure 3.4).

3.6.3 AcrD

AcrD is not encoded with its own PAP and instead forms a tripartite complex with AcrA and TolC in the same manner as AcrB. As seen for AcrB the K-12 protein sequence of AcrD was highly conserved across the *E. coli* STs however a single substitution was highly conserved at position 656 of AcrD where a glutamine had been substituted for a lysine in 100% of the B2, D, and F assemblies but also 100% ST3 and ST17 assemblies which are both B1 (Figure 3.5).

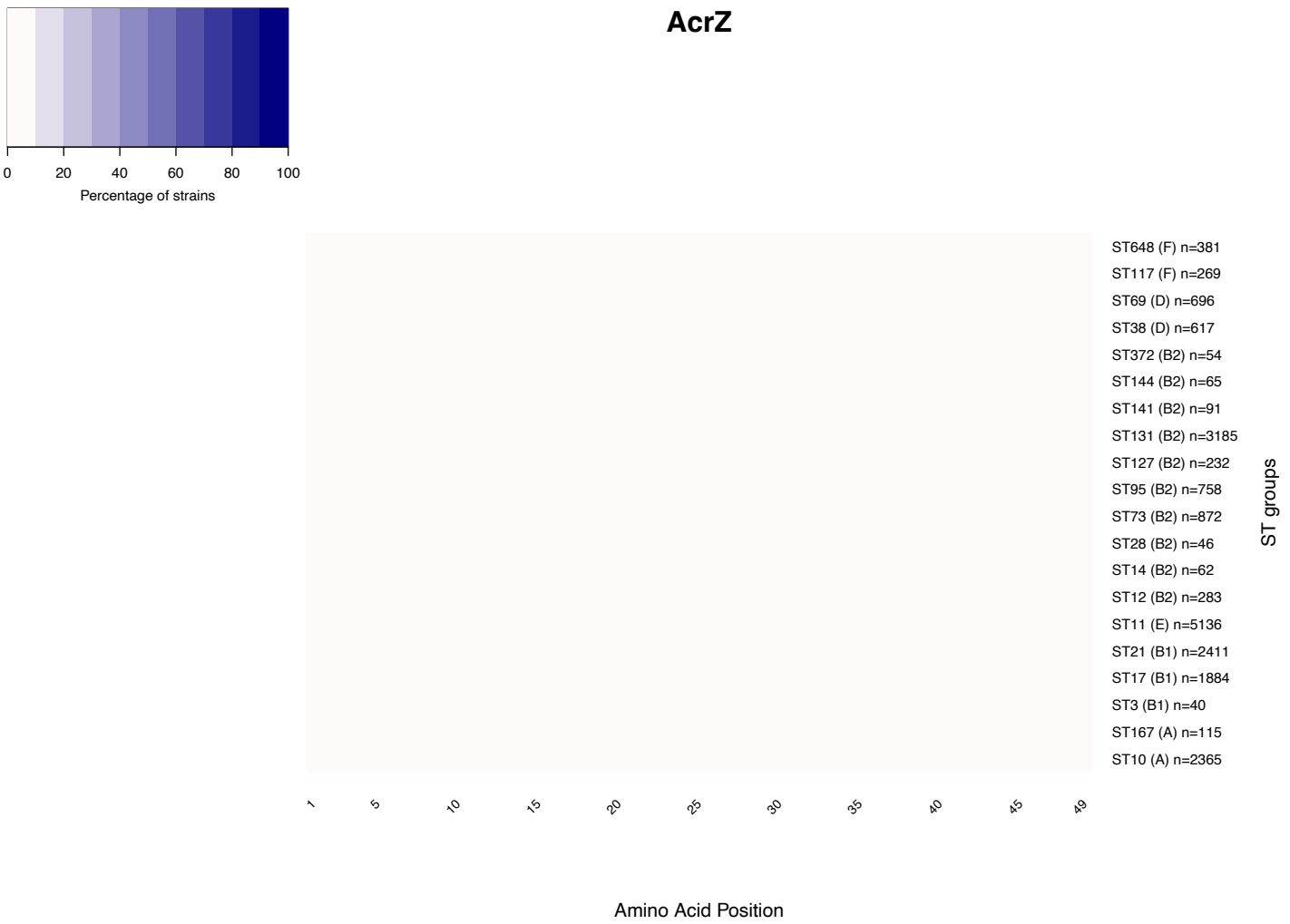


Figure 3.4 Conservation of the AcrZ protein across 20 *E. coli* STs

The conservation of AcrZ proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to AcrZ (NP_415283.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. *N* value shows the number of protein sequences assemblies identified and analysed.

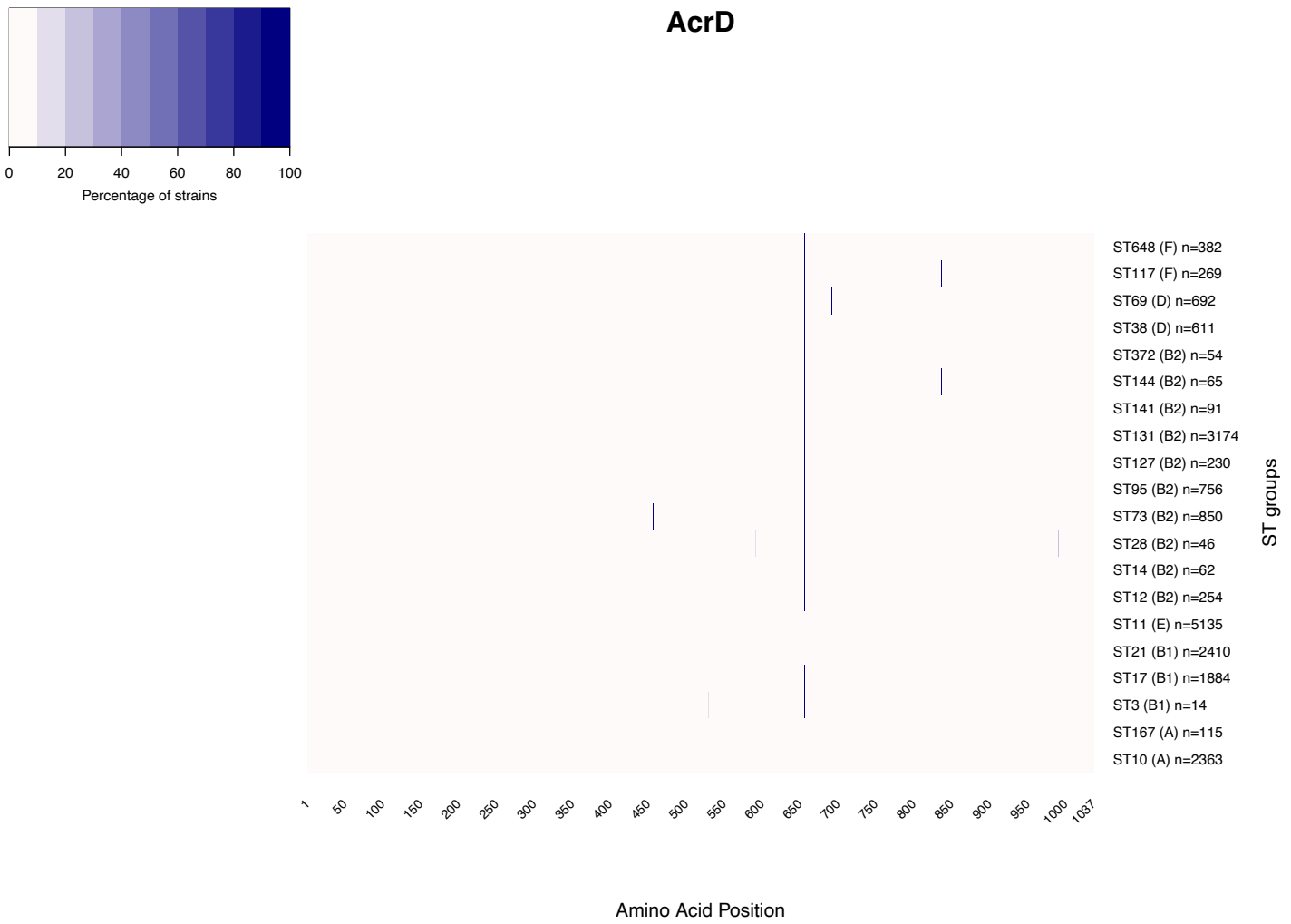


Figure 3.5 Conservation of the AcrD protein across 20 *E. coli* STs

The conservation of AcrD proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to AcrD (NP_416965.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

3.6.4 AcrEF

AcrE is the PAP associated with AcrF and TolC, an efflux system that has been shown to affect virulence in *Salmonella* (Nishino *et al.*, 2006). Across the STs, the K-12 AcrE peptide sequence was found to be more varied than AcrA as lineage specific amino acid substitutions were conserved across multiple regions of the protein (Figure 3.6). Across phylogroups B2, D and F, an N103S substitution was seen in 100% of the assemblies. A second highly conserved variation was seen at position 327 where instead of the K-12 aspartic acid phylogroups B2 lineages (except ST144) and ST117 (F) had an asparagine.

Across the 20 *E. coli* ST in the analysis, the AcrF proteins were found to have much greater variation from the K-12 sequence than the homolog AcrB (Figure 3.7). An F240L substitution was highly conserved and found in 100% of the B1, E, B2, D and F assemblies. Also at position 770 the K-12 reference sequence was found to possess a leucine, however across 100% of the phylogroups B1, B2, D and F assemblies this had been substituted for valine. Assemblies belonging to B2 were also found to have highly conserved substitutions at positions 338 (H338Q), 428 (K428R) and 637 (N637S). While S806A and K849Q were exclusively found in phylogroups D and F, notably these two substitutions were present in 99.6% of the assemblies (mean) belonging to those phylogroups (D and F). ST specific substitutions were seen across ST11, however only 29 peptide sequences were included in the analysis due to the strict cut-off value used for filtering BLAST results (protein length). Briefly ST11 possessed a gene insertion resulting in a truncation midway through the protein resulting in the low number of sequences identified. This will be further expanded upon in Chapter 4.

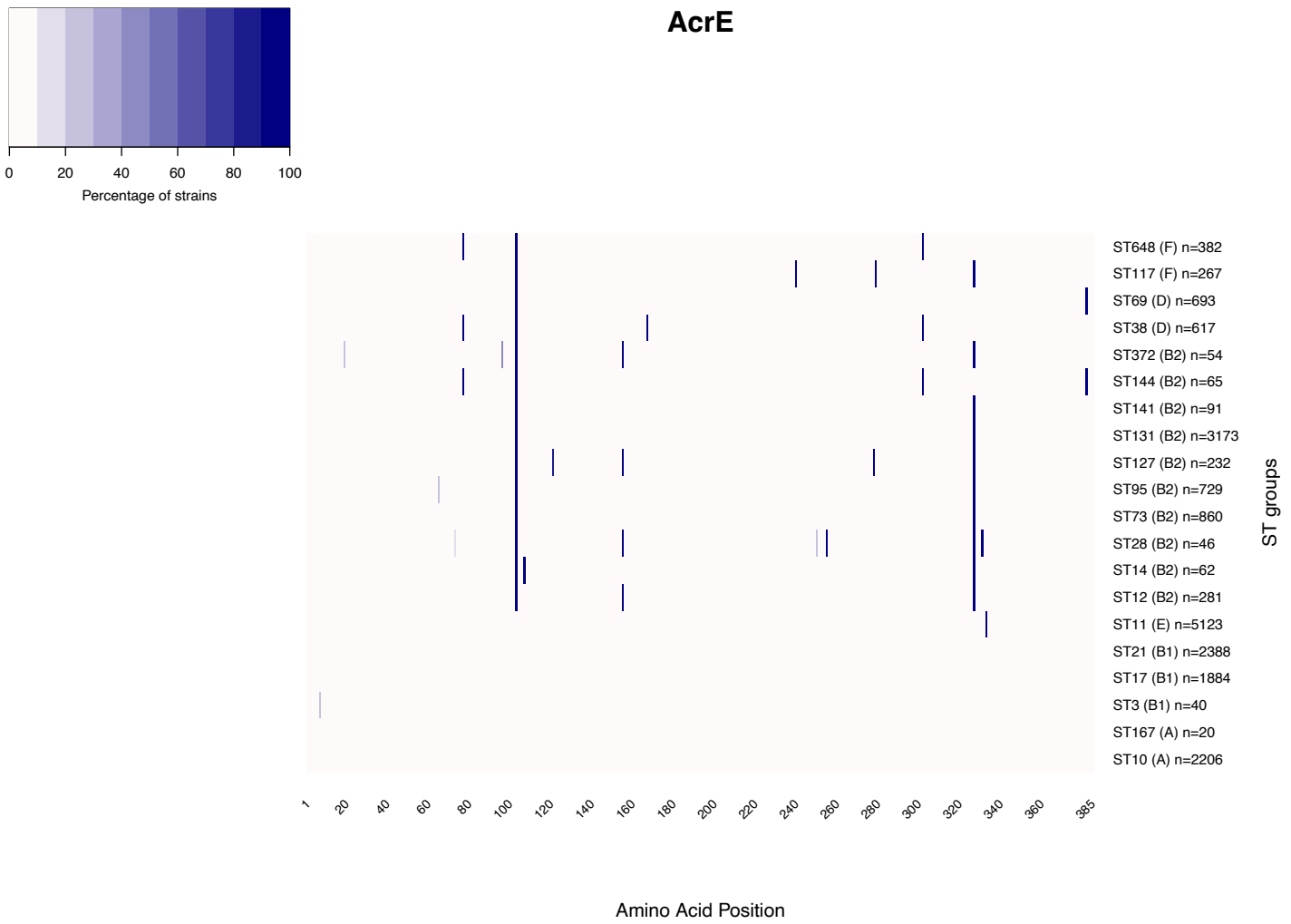


Figure 3.6 Conservation of the AcrE protein across 20 *E. coli* STs

The conservation of AcrE proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to AcrE (NP_417731.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

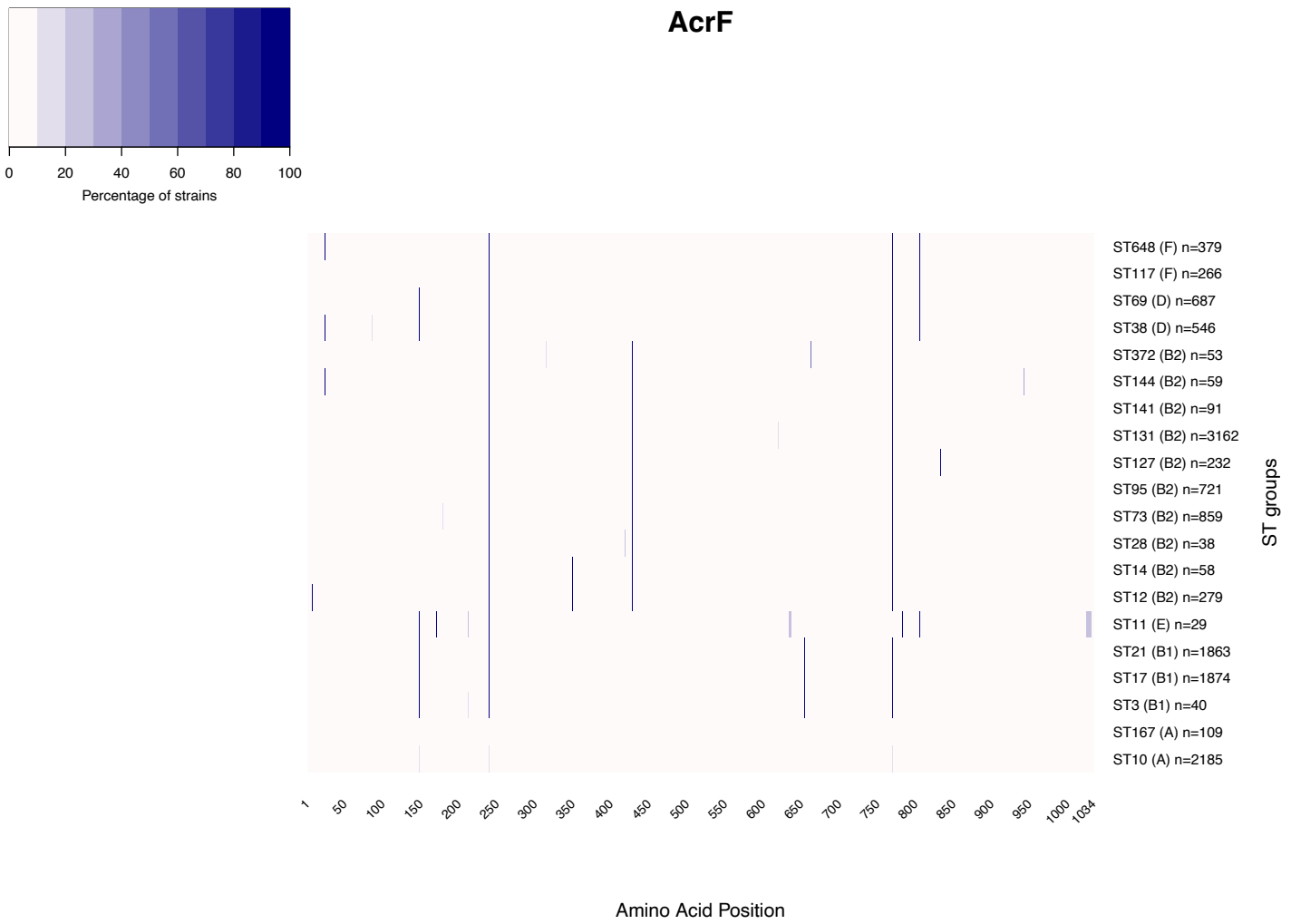


Figure 3.7 Conservation of the AcrF protein across 20 *E. coli* STs

The conservation of AcrF proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to AcrF (NP_417732.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

3.6.5 CusABC

CusA is the only RND pump in *E. coli* that does not use the OMP TolC. Furthermore CusABC has the narrowest substrate range of the RND pumps. The CusA peptide sequence present across *E. coli* was found to be broadly conserved but with five substitution hot spots (Figure 3.8). These I267V, E634D, A695T, I804V and V955E substitutions were found to be highly conserved across the B2 phylogroup, however at each position ST372 (phylogroup B2) had maintained the K-12 I267, E634, A695, I804 and V955.The V955E substitution was also absent in ST12 (B2).

CusB also had substitutions present across the phylogroup B2 assemblies, however at amino acid 25 the K-12 valine was only present in phylogroup A. In every other phylogroup the valine had been replaced by a phenylalanine (Figure 3.9). B2 specific substitutions were present at positions 42 and 107. An I42V substitution was identified in 100% of B2 assemblies included in the analysis, while at 107 STs 28 and 95 had maintained the K-12 alanine but in STs 12, 14, 73, 127, 131, 141, 144 and 372 a serine was found in over 98.6% assemblies (mean). Additional changes in relation to K-12 were seen at amino acids 148 and 337. A T148A substitution was found across phylogroups B2 and E, and S337N was present across B2, D and ST648 (F).

Across the sequence of CusC, B2 assemblies once again had positions in which the K-12 amino acid was absent and alternative amino acids were highly conserved. T81A, D117N and T118S were present in 100% of the B2 assemblies, the D117N and T118S substitutions were also present in phylogroups E, D and ST117 of phylogroup F (Figure 3.10). A fourth variant was seen at position 273, here ST14 and ST131, both of which are phylogroups B2, had the K-12 amino acid of serine but every other B2 lineages had a proline at this position. As seen in AcrF, lineage specific substitutions

could be seen across ST14, across the 62 assemblies L124F, F152L, V342I, V370I, T386S and A389D were present in 100% of the genomes.

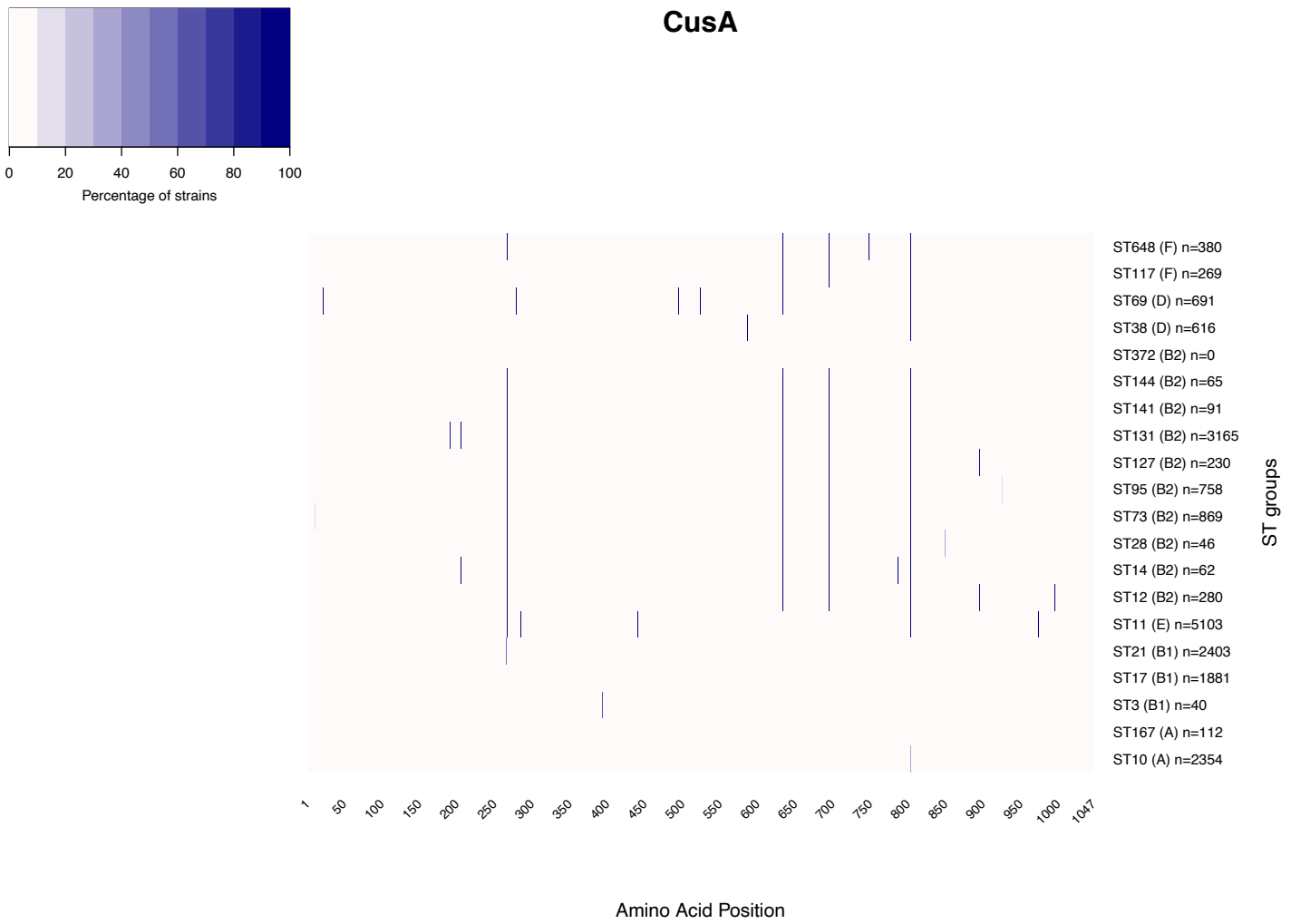


Figure 3.8 Conservation of the CusA protein across 20 *E. coli* STs

The conservation of CusA proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to CusA (NP_415107.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

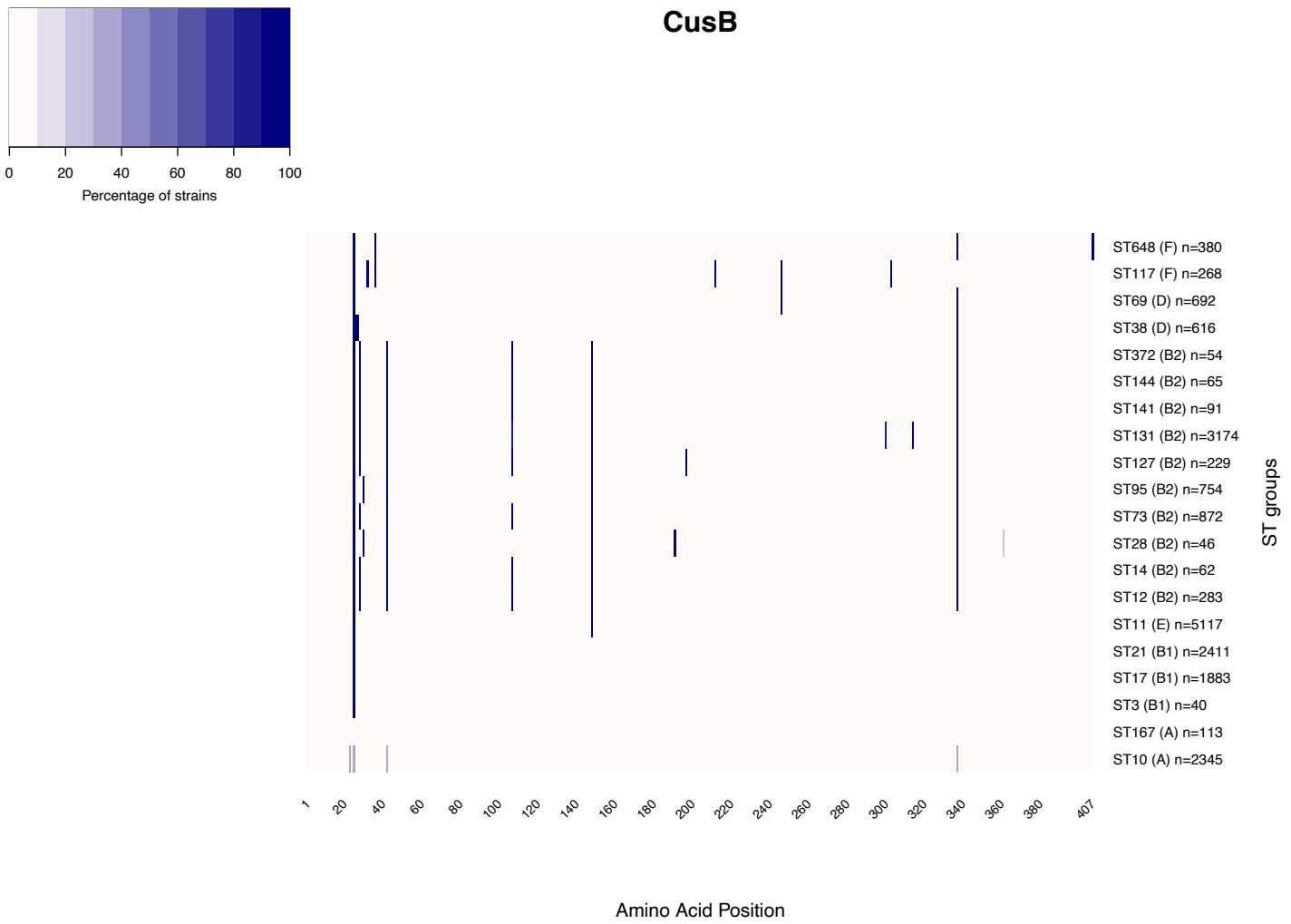


Figure 3.9 Conservation of the CusB protein across 20 *E. coli* STs

The conservation of CusB proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to CusB (NP_415106.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

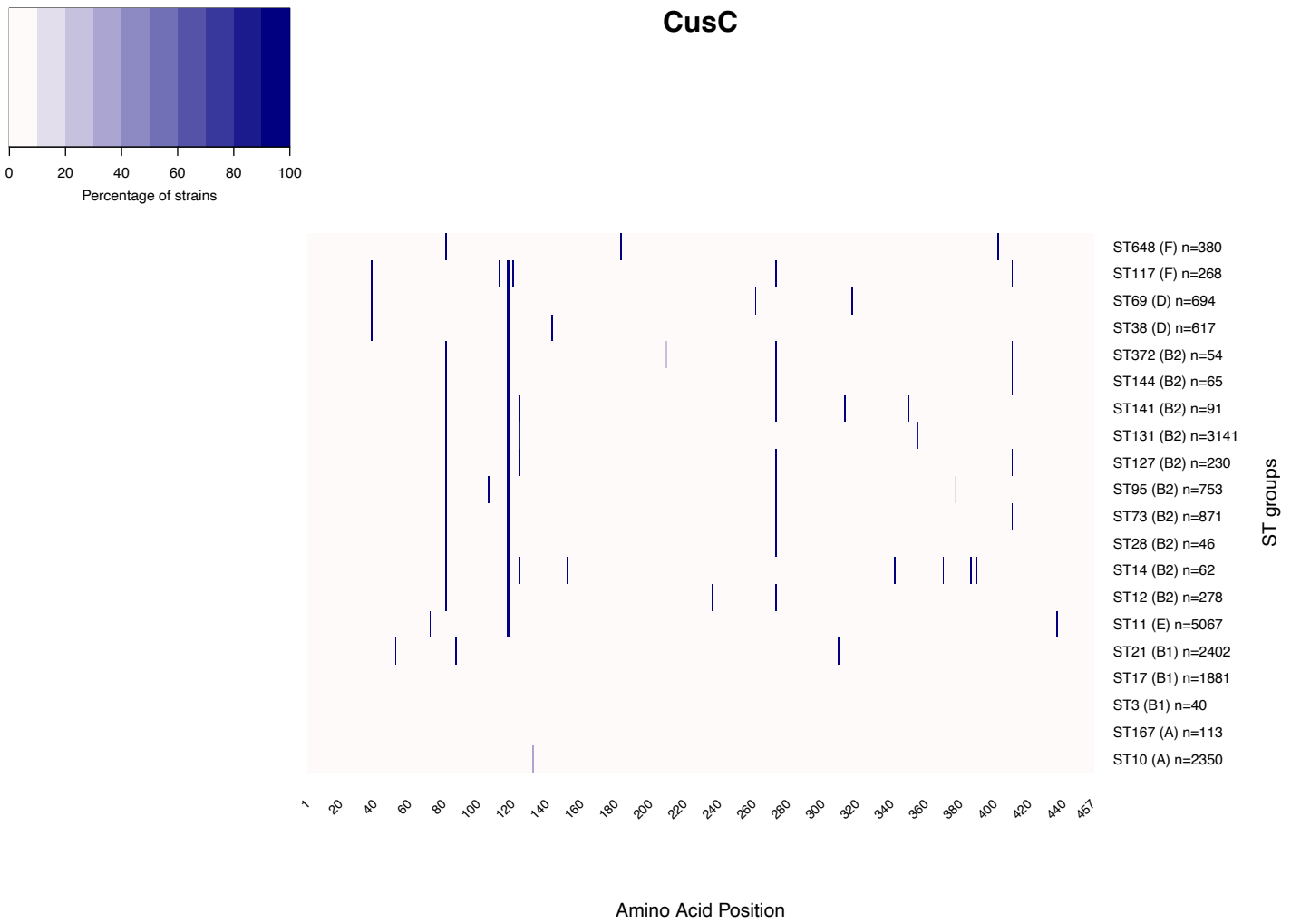


Figure 3.10 Conservation of the CusC protein across 20 *E. coli* STs

The conservation of CusC proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to CusC (NP_415104.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

3.6.6 MdtABC

The MdtABC complex highlighted that the amino acid sequences of K-12 that not only differed from those belonging to other phylogroups, but also to assemblies within phylogroup A in which *E. coli* K-12 belongs. Across MdtA R33Q was present in all B2 and E assemblies and also highly conserved in ST3 and ST17 which belong to phylogroup B1 (Figure 3.11). A B2-specific substitution was identified at position 56, where the K-12 serine had been substituted to alanine. However, at position 99, an isoleucine to methionine substitution was present across all phylogroups, though less prevalent in phylogroup D. While the I99M change is conservative, it highlights that the K-12 sequence is not representative of the wider population.

MdtB is a 1040 amino acid protein that together with MdtC forms the only heterotrimer RND pump in *E. coli.* There is conflicting evidence as to whether either of the units can form a functional RND protein without the other (Kim *et al.*, 2010, Wang and Fierke, 2013). The K-12 MdtB sequence was highly conserved across *E. coli,* a T212S substitution was seen in STs 14, 127, 131, 141 and 144, all of which belong to phylogroup B2 (Figure 3.12). A L910M change was also prevalent across phylogroups B2, F, ST38 (E) and although at low levels, present in in phylogroups A and B1 sequences.

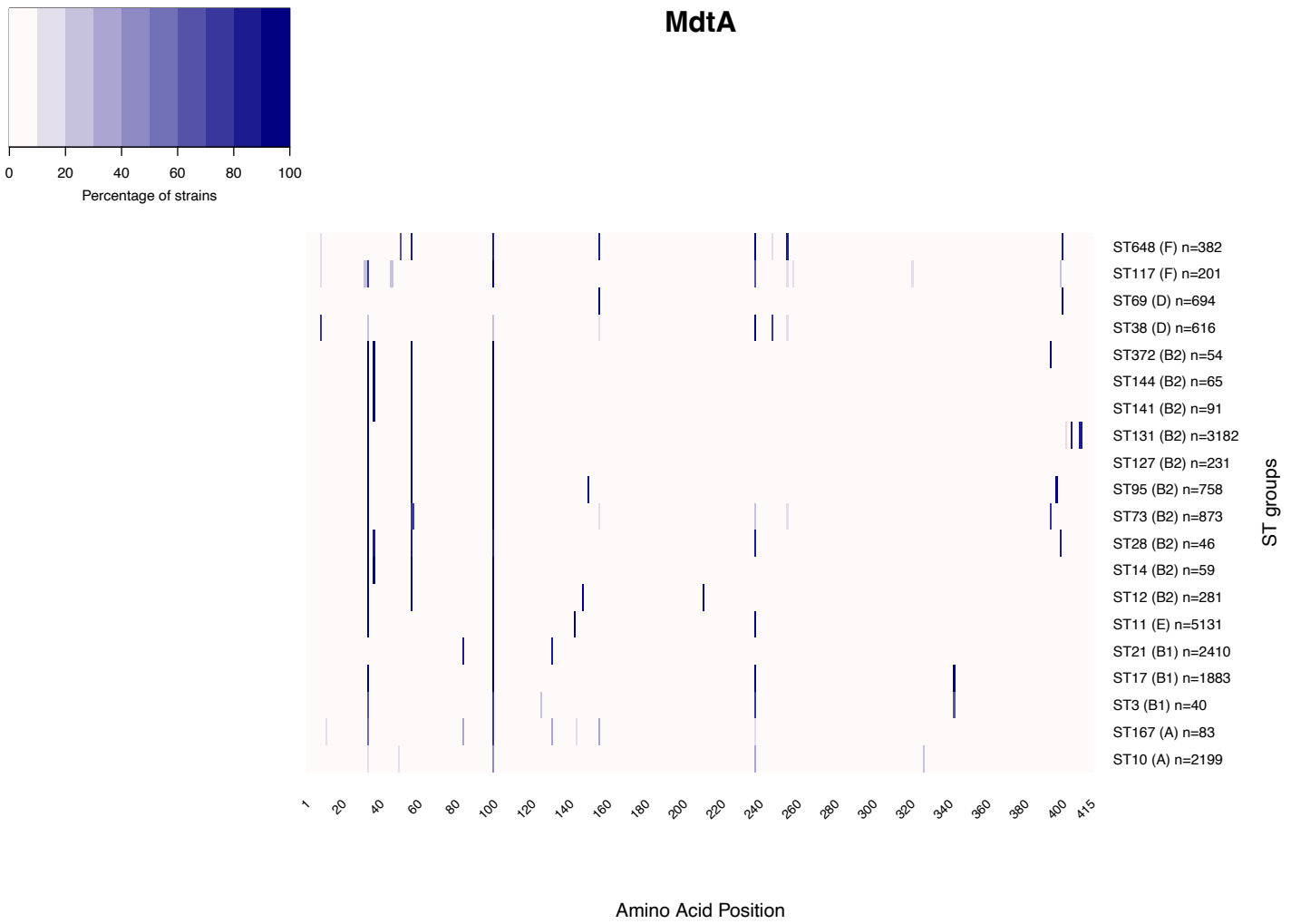


Figure 3.11 Conservation of the MdtA protein across 20 *E. coli* STs

The conservation of MdtA proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to MdtA (NP_416578.2) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

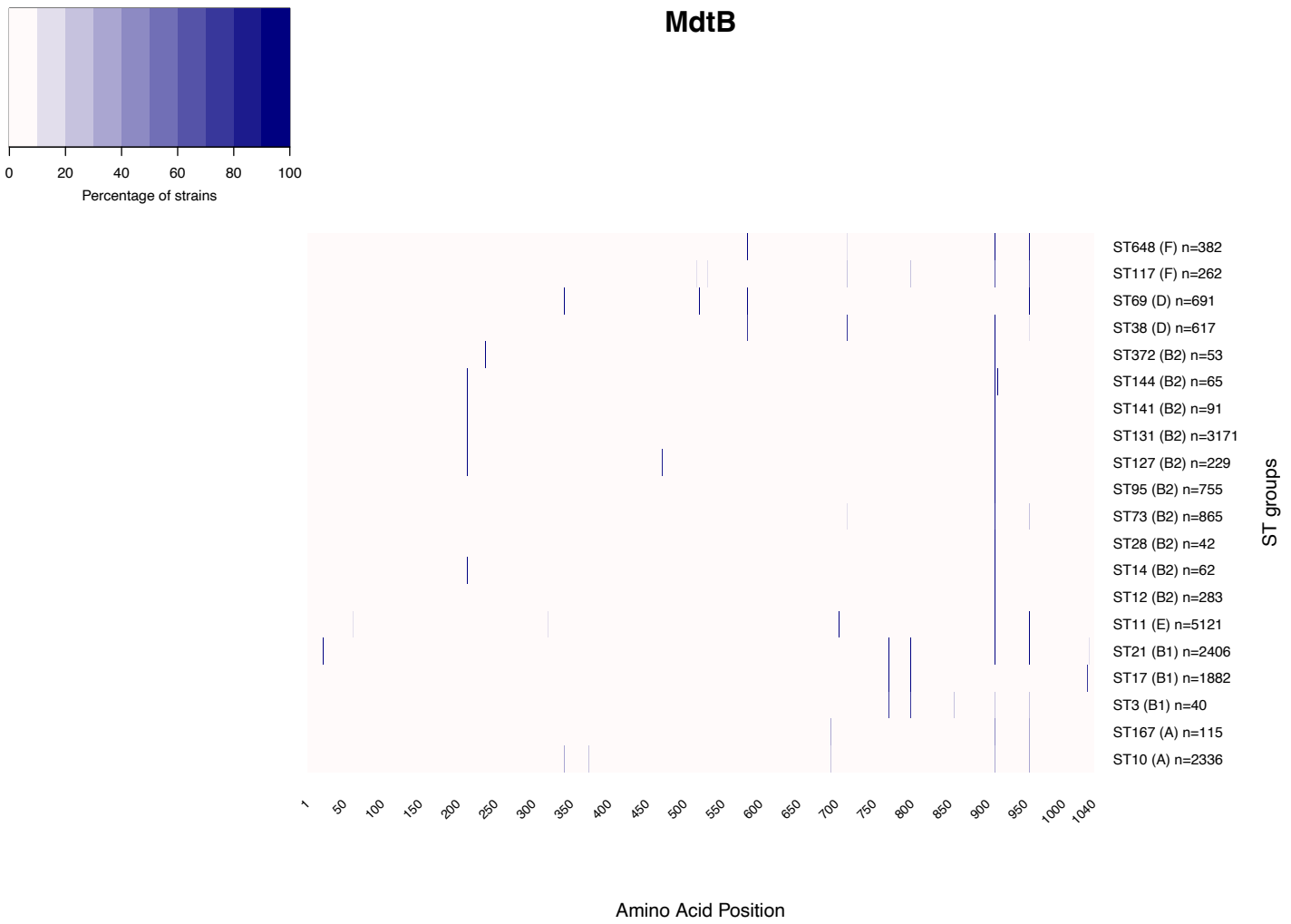


Figure 3.12 Conservation of the MdtB protein across 20 *E. coli* STs

The conservation of MdtB proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to MdtB (NP_416579.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

MdtC was far less conserved than MdtB. Across the 20 STs, the alternative translational start codon valine was present in MdtC with only the K-12 sequence found to encode the methionine start codon, furthermore at 46 and 208, I46M and V208I substitutions were seen across all phylogroups (Figure 3.13). The conservation of I46M and V208I was above 93.4% across B1, B2, D, E and F genome assemblies, and in phylogroup A in which K-12 belongs, I46M was in 62.1% of ST10 and 100% of ST167, while V208I in 43.9% and 61.7% of ST10 and ST167 respectively. Additional substitutions were seen across the protein, predominantly present across the B2 phylogroup. These were K303R, E309S, S365A, D617G, V632K, S657A, Q698E, V707I and T1022A.

3.6.7 MdtEF

Less sequence diversity was seen across MdtEF than MdtABC and CusABC. The PAP MdtE was found to have minimal variation across the protein when compared with K-12. However, again there were specific positions where variation occurred and is conserved across a lineage and sometimes a phylogroup (Figure 3.14). Though the MdtE protein sequences present across the 20 ST where highly conserved, a S251G change was present in 100% of phylogroup B2 genome assemblies and a S107T change was found exclusively in D and F assemblies.

Like in MdtE, there are amino acid substitutions that were highly conserved in MdtF across phylogroup B2 assemblies. Both D602N and I959V were found only in B2 assemblies, however these substitutions were present in 100% of assemblies (Figure 3.15). In addition to these changes, A295T was present in 100% of B1, E, B2 and ST69 assemblies. E334K was found across B1, E and B2, F and ST69 assemblies. A fifth

substitution was present across a range of STs at 712, here the K-12 threonine was an isoleucine. This substitution was present in 100% of phylogroup B2 and 98.6% of F.

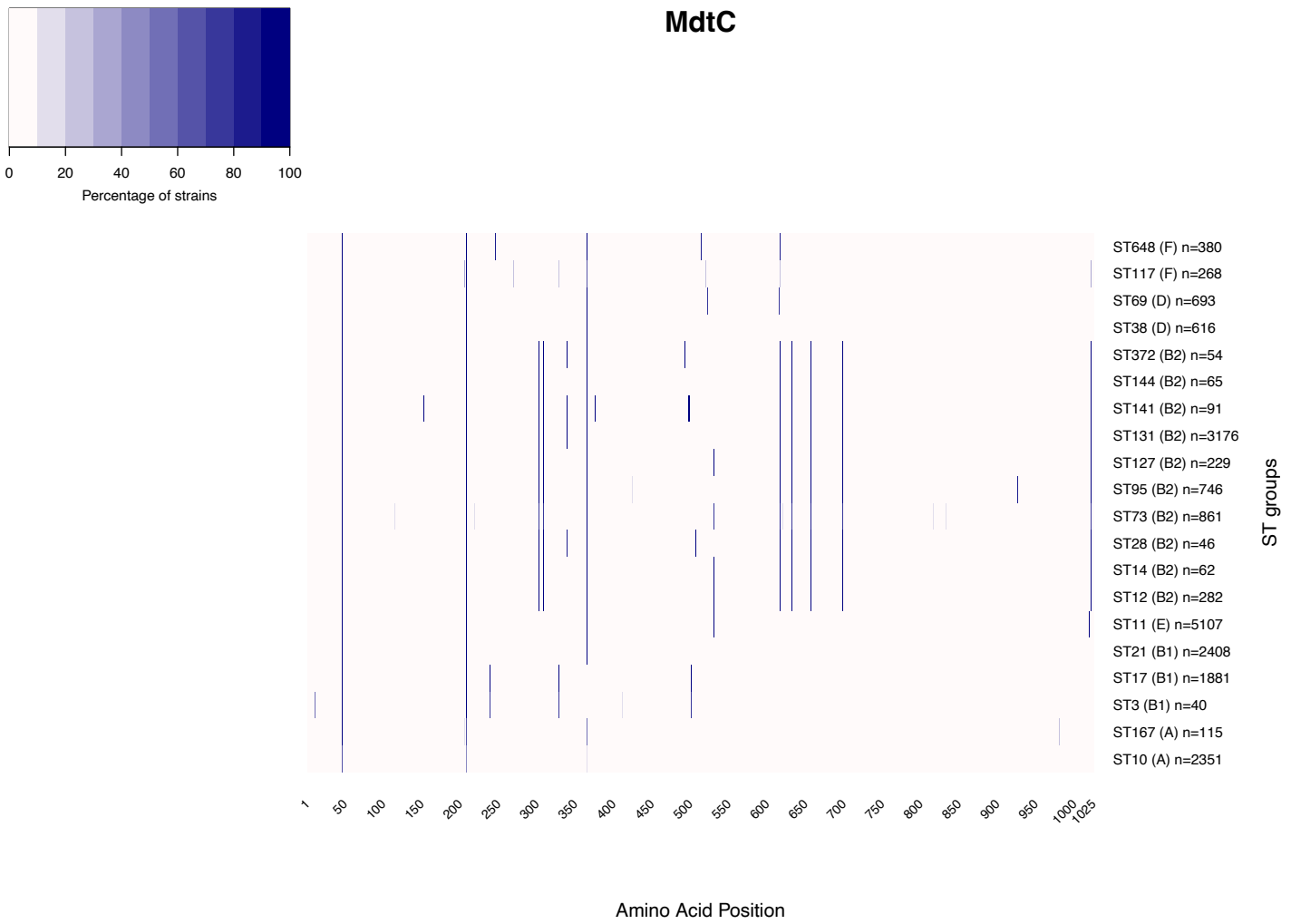


Figure 3.13 Conservation of the MdtC protein across 20 *E. coli* STs

The conservation of MdtC proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to MdtC (NP_416580.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

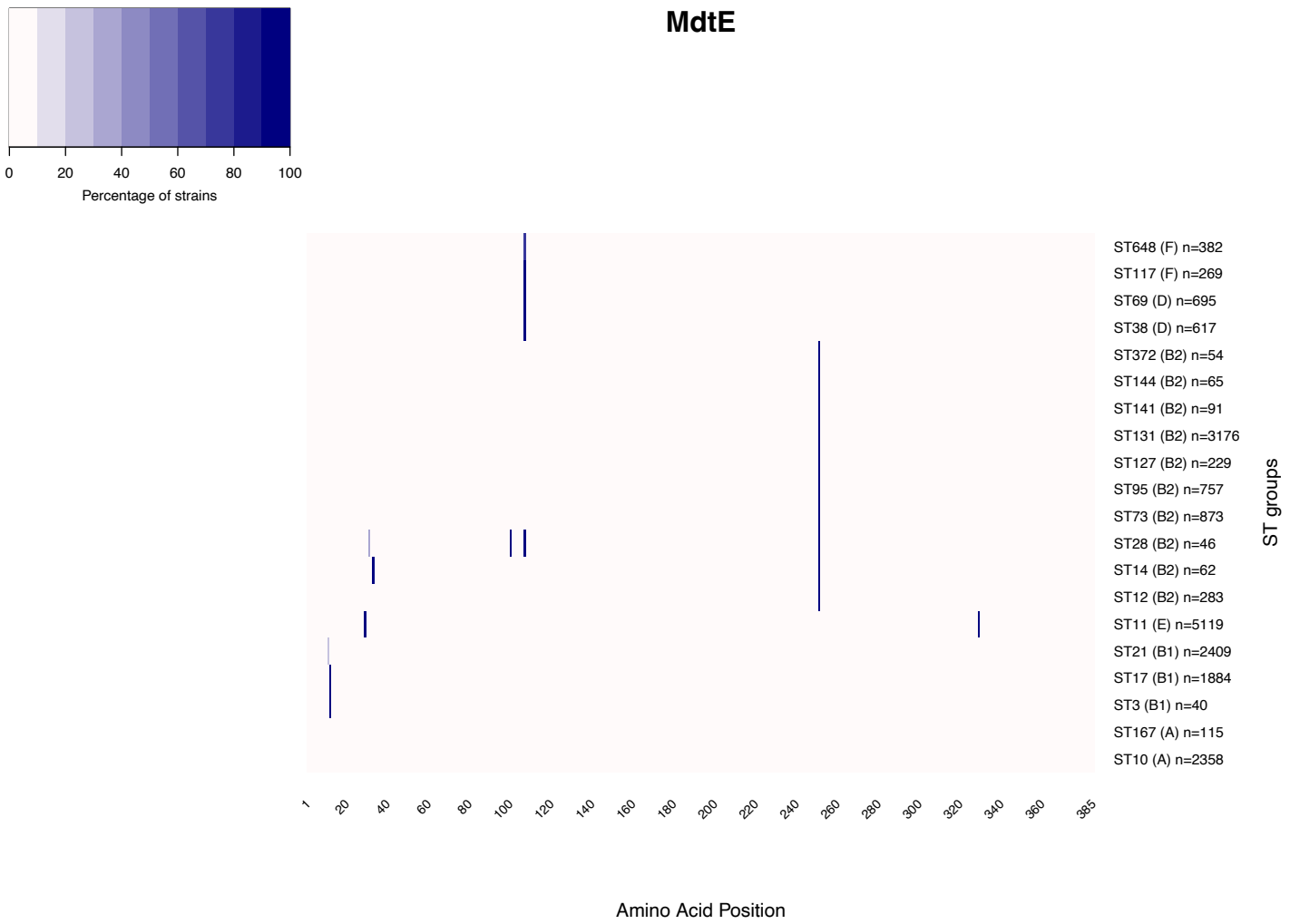


Figure 3.14 Conservation of the MdtE protein across 20 *E. coli* STs

The conservation of MdtE proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to MdtE (NP_417970.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belonging to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

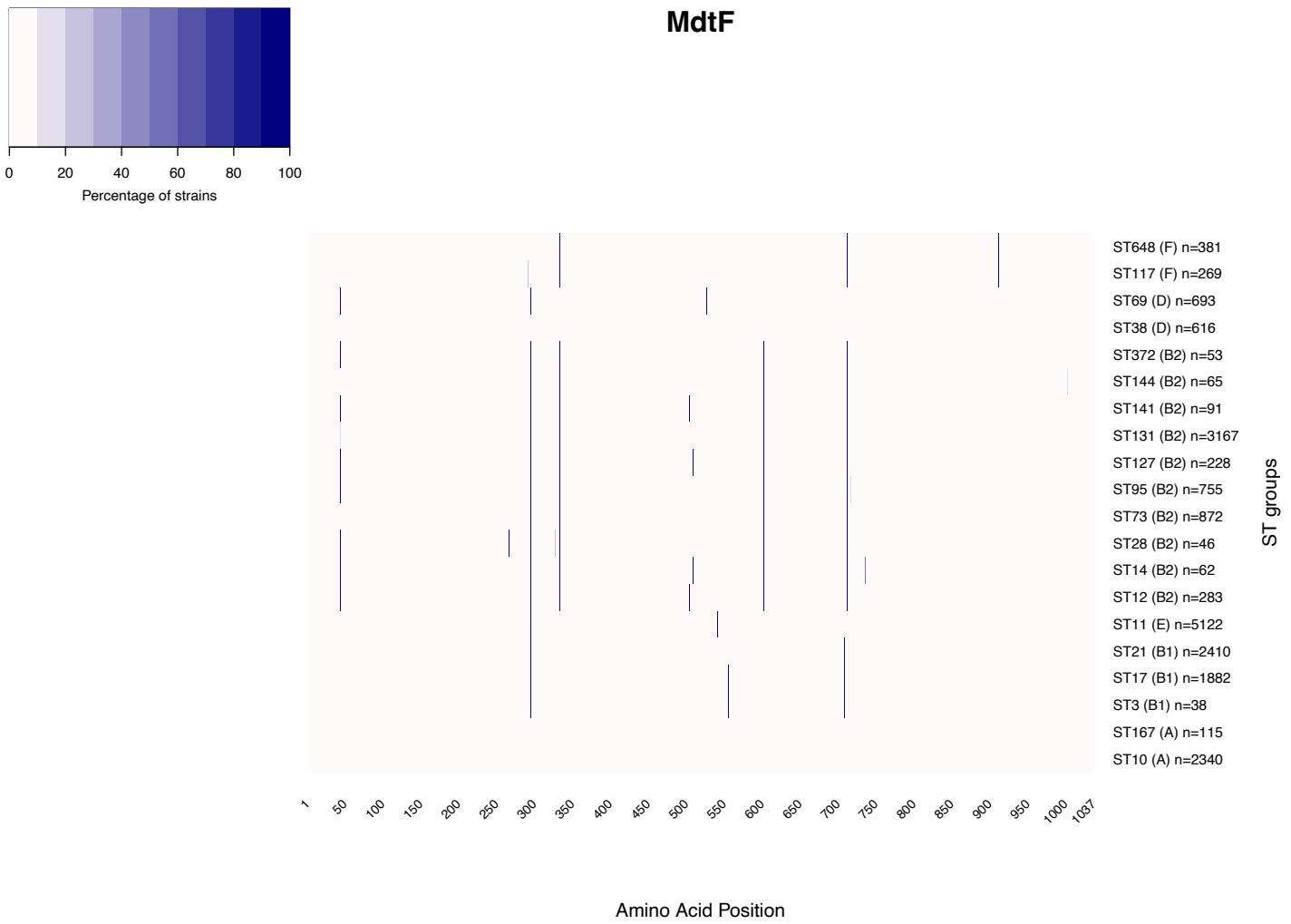


Figure 3.15 Conservation of the MdtF protein across 20 *E. coli* STs

The conservation of MdtF proteins across the STs is shown above as a heatmap. Each assembly from all 20 *E. coli* STs was aligned to MdtF (NP_417971.1) from *E. coli* K-12 MG1655 (NC_000913.3). Where the K-12 amino acid is present in 100% of the assemblies belong to an ST, the cell is white. Where an amino acid differs in an ST, the cell is blue. The deeper the shade of blue, the greater the percentage of assemblies possess an alternative amino acid to K-12 at the given position. *N* value shows the number of protein sequences assemblies identified and analysed.

3.7 Discussion

RND efflux pumps are highly important for the survival of Gram-negative bacteria, allowing colonisation and replication in harsh conditions such as extreme pH, in the presence of human host molecules such as bile salts and indole, metals such as copper and silver, and antimicrobial drugs and disinfectants (Nishino *et al.*, 2003, Nishino and Yamaguchi, 2001, López *et al.*, 2017, Ma *et al.*, 1995, Schuster *et al.*, 2021, Novoa and Conroy-Ben, 2019). The data presented in this chapter demonstrates that RND efflux pumps are highly conserved, however the level of conservation differs between RND systems.

The best studied RND pump in *E. coli* is AcrB. This has the broadest substrate range which includes clinically relevant antimicrobials. While its full natural role is not completely understood, the structure of AcrB and its role in MDR is well studied and studies of highly homologous systems suggest it may also be involved in virulence and biofilm production. Given the clear biological importance of AcrB it was not surprising to find that *acrB* is highly conserved across the 20 ST pangenomes investigated. Furthermore, AcrA, the native PAP of AcrB was found to be equally highly conserved with a mean of 1.2 alleles across the pangenomes. This is probably predominantly because it supports efflux through the major pump AcrB. However, AcrA is additionally important because it is also the PAP required for function of the AcrD efflux pump and can also form promiscuous interactions with AcrF (McNeil *et al.*, 2019).

Across the six *E. coli* RND systems, five form a tripartite complex with the OMP TolC, therefore high conservation across the 20 STs was expected. Only two STs were found to have more than one *tolC* allele, these were ST10 and ST117, both of which had two alleles 1482 nucleotides long. Inactivation of *tolC* has been shown to not only

effect the MDR phenotype of *E. coli* but also the export of bacterial toxins, cell metabolism and pH tolerance (Deininger *et al.*, 2011, Dhamdhere and Zgurskaya, 2010, Delgado *et al.*, 1999). The high conservation of *tolC* at the pangenome level is therefore likely to have occurred as changes in nucleotide length would result in significant alterations in protein sequence and potentially structure. This in turn would have considerable effects on all *E. coli* RND systems except CusABC.

At the protein level, both AcrA and AcrB were highly conserved with little variation from the *E. coli* K-12 sequence seen across the 20 STs. However in both proteins a single amino acid substitution was present and highly conserved across each protein. In AcrA, the K-12 threonine had been replaced by an alanine in >99.9% of assemblies belonging to B2, D and F. The H596N AcrB substitution was present across phylogroups B2 and D, however the K-12 histidine was maintained in phylogroups F and selected B2 STs (ST12 and ST95). In TolC an A233T substitution was highly conserved again across phylogroup B2, though only present in one of the phylogroup D and F lineages. These three amino acid substitutions will be further characterised in the Chapter 5.

The most conserved gene across the pangenome was *acrZ*, an accessory peptide shown to modulate AcrB substrates (Du *et al.*, 2020). The *acrZ* gene encodes a 49 amino acid protein shown to modulate the substrate specificity of AcrB with an interaction that results in a conformational change of the binding pocket (Hobbs *et al.*, 2012, Du *et al.*, 2020). Though its function is not fully understood, mutations that impact the helix of the peptide have been shown to impact cell growth in the presence of chloramphenicol (Du *et al.*, 2020). The *acrZ* gene was the only nucleotide sequence to have no more than one allele per ST, suggesting a biological importance for the 150

nucleotide sequence. Furthermore, no significant substitutions were seen across the protein sequences at the genome assembly level suggesting the sequence present in K-12 is under high selection.

Similar to AcrB, AcrD forms a tripartite complex with AcrA and TolC however its substrate range differs. Across the pangenomes of 20 ST, the mean number of alleles for *acrD* was 1.3 however none were identified for ST3, suggesting the gene was not present in this ST. This was not found to be the case when determining the protein conservation, however only 14 of the 40 assemblies in the analysis possessed a protein of the expected length (1034 nucleotides). Despite the limited number of AcrD sequences identified across the ST3 assemblies, an amino acid change was identified at 656. This Q656K was found in 100% of ST3 and ST17 assemblies, both of which belong to phylogroup B1, and 100% of B2, D and F assemblies. Glutamine was only present at position 656 in the reference sequence of K-12, phylogroups A and E. Amino acid 656 is located in loop 2 of AcrD, a region of the protein involved in substrate recognition (Elkins and Nikaido, 2002). A glutamine to lysine change may therefore have an effect on substrate recognition.

The most varied genes across the pangenomes were *acrF* and *mdtC,* both of these genes were found to have an average of 1.6 alleles. Despite being one of the most variable genes across the pangenomes, across both the ST38 (D) and ST11 (E) pangenomes no *acrF* allele was identified. In ST38, analysis using the protein sequence identified an AcrF protein sequence in 546 of the 617 assemblies used for the pangenome construction. However, when identifying AcrF using TBLASTN only 29 of the 5137 ST11 assemblies possessed a match, suggesting this RND pump was absent from ST11. This finding is further explored in Chapter 4. However, protein

sequences across ST11 which had the expected length of 1,034 amino acids were found to have significantly more variable amino acids in relation to K-12 than seen in AcrB. Suggesting that although being a homolog of AcrB and having significant substrate overlap with this RND pump, the protein sequence of AcrF is of less importance. As the *acrEF* operon has been shown to be H-NS silenced in both *E. coli* and *S.* Typhimurium this suggests that its expression is not of critical importance in most circumstances (Nishino *et al.*, 2009, Nishino and Yamaguchi, 2004).

MdtBC is the only heterotrimer RND pump in *E. coli* and has a stoichiometry of B_2C (Kim *et al.*, 2010). It is believed the two-protein RND pump arose through a gene duplication and the two proteins have subsequently adapted to function together (Gorecki and McEvoy, 2020), though is unclear whether either subunit can function as a homotrimer (Wang and Fierke, 2013, Fricke *et al.*, 2008). At the pangenome level, the nucleotide sequence of *mdtB* was more conserved than that of *mdtC* suggesting MdtB biologically more important than MdtC. This was again seen in the protein analysis as throughout the MdtC protein, amino acid substitutions with respect to the K-12 sequence had occurred throughout the different phylogroups but more specifically B2. It is worth noting that while K-12 possessed a methionine start codon, 100% of genome assemblies in ST10 of which K-12 belongs possessed an alternative start codon, valine. However this would not alter the first amino acid in the peptide sequence. This variation of start codon was seen across all STs, suggesting K-12 is the anomaly. It has been proposed that MdtB and MdtC perform different functions during the functional rotating mechanism, with MdtB enabling the conformational change required for proton transport and MdtC involved in substrate binding and transport (Kim and Nikaido, 2012). Due to their differing functions it is possible that one

of the protein monomers can vary more so than the other without having a deleterious impact on function.

Recent work has begun to elucidate the role of MdtEF, a PAP and RND pump co-located with the *gadE* operon. GadE is a key regulatory component of acid resistance in *E. coli* and crucial for ST11 survival in the simulated gastric environment (Kailasan Vanaja *et al.*, 2009, Deng *et al.*, 2013). Upregulation of this operon has been reported at the transition from exponential to stationary phase and during anaerobic growth (Kobayashi *et al.*, 2006). Furthermore deletion of *mdtEF* increases *E. coli* susceptibility to indole nitrosative derivatives (Zhang *et al.*, 2011, Deng *et al.*, 2013). More recently, deletion of *mdtE* has been shown to increase strain fitness, however following a 2 hour exposure to extreme pH conditions that are below what supports growth, MdtEF-TolC confers a fitness advantage (Schaffner *et al.*, 2021). High conservation of this system was seen at both the pangenome and protein level, suggesting high selection pressure. In MdtE the S251G variant was present and extremely highly conserved, the same was seen for both MdtF D602N and N709D. These substitutions were conserved across the STs in phylogroup B2. Taken together with recent functional characterisation, it is highly like that MdtEF-TolC is important in oxygen scarce and high pH environments, such as those encountered when causing infection.

The greatest variation of RND genes was seen across both ST10 and ST131 which were also the two largest pangenomes in previous work by CC (Connor, 2021). *E. coli* isolates belonging to ST10 are part of phylogroup A and are isolated from both humans and animal intestinal samples as both commensals and pathogens (Manges and Johnson, 2012). While this lineage is most commonly associated with drug

susceptibility, ST10 is also a reservoir of *mcr-1* (Matamoros *et al.*, 2017). Carriage of AMR genes also occurs across ST131 which is a clonal complex associated with ExPEC infections (Oteo *et al.*, 2009, Garcia-Menino *et al.*, 2018). ST131 has a small core genome, despite this variation across the accessory genome in seen due to the presence of plasmid and virulence factors that vary across the lineage (Petty *et al.*, 2014). High AMR carriage across these two lineages is likely due to antimicrobial exposure, while mechanisms reported across ST10 and ST131 are mainly enzymatic, variation in RND pump sequences may be selected for due to subsequent effects these sequence changes have on efflux function or efflux pump complex stability.

In all RND-associated proteins except AcrZ amino acid substitutions were highly conserved across the phylogroup B2. B2 is associated with AMR and pathogenesis, furthermore human isolates of phylogroup B2 possess a higher number of virulence factors than non-human isolates (Touchon *et al.*, 2020). As both the lifestyle and RND peptide sequences differ between *E. coli* K-12 and phylogroup B2, data suggests K-12 is a poor model for determining RND structures across *E. coli* generally but more specifically for phylogroup B2 isolates.

3.8 Key findings

- RND genes in *E. coli* are generally highly conserved.
- The most highly conserved RND genes were *acrA, acrB, tolC, acrZ, mdtE* and *mdtF.*
- The most variable RND genes were *acrF* and *mdtC.*
- STs ST10 and ST131 had the greatest variation in the sequence of the RND genes across the 20 ST pangenomes.

- In AcrA, AcrB and TolC single amino acid substitutions were conserved across phylogroups B2, D and F. These will be explored further in Chapter 5.

3.9 Future work

- Introduce amino acid substitutions to the sequence of AcrZ and phenotypically characterise the implications on AcrAB-TolC function with MIC, biofilm production and virulence assays using the *Galleria mellonella* model for preliminary experiments
- Characterise MdtE S251G, MdtF D602N and MdtF T712I due to their high conservation across phylogroup B2
- Determine the significance of MdtB conservation.
- Expand this experimental approach to further *E. coli* STs to determine if amino acid polymorphisms at identified amino acid positions are present across the wider population of *E. coli.*
- Similarly this approach could be applied to other organisms in which efflux is a contributing factor to MDR

Chapter 4 The *E. coli* lineage ST11 does not possess a functional AcrF protein

4.1 Background

E. coli ST11 (O157:H7) is a frequent foodborne human pathogen that causes symptoms including stomach cramps and bloody diarrhoea. The primary ST11 niche is the recto-anal junction in cattle where the organism reproduces asymptomatically (Naylor *et al.*, 2003, Pruimboom-Brees *et al.*, 2000). In 2017, 2,363 cases were reported in the United States, while in 2018, 607 cases occurred in England and Wales (CDC, 2021, PHE, 2020). In severe cases, *E. coli* ST11 infection can result in HUS, an ST11-assoicated form of kidney failure due to its carriage and expression of Stx (Sauder and Kendall, 2018). Hospitalisation is associated with ST11 infection. A 2019 US outbreak of O157:H7 caused by contaminated lettuce resulted in 167 infections across multiple states, 15 of these individuals went onto develop HUS (CDC, 2020). The following year, a UK outbreak resulted in 36 infections with over a third of people requiring hospital treatment (Mulchandani *et al.*, 2021).

AcrF is a close homolog of AcrB in *E. coli* and both share 77% of their amino acid identities (Nishino and Yamaguchi, 2001). There is also wide overlap in the substrate ranges of AcrF and AcrB (Nishino and Yamaguchi, 2001). In addition to AMR, RND efflux pumps have been shown to be involved in virulence across Gram-negative bacteria, including *Salmonella, Klebsiella pneumonia, Enterobacter cloacae* and *Erwinia amylovora* (Burse *et al.*, 2004, Padilla *et al.*, 2010, Wang-Kan *et al.*, 2017, Pérez *et al.*, 2012). While AcrB is expressed constitutively, under laboratory conditions the *acrEF* operon is expressed at very low levels due to HNS silencing but like AcrB,

absence of AcrF has been shown to attenuate virulence (Nishino and Yamaguchi, 2004, Ma *et al.*, 1995, Nishino *et al.*, 2006, Blair *et al.*, 2015b). Due to the virulence of ST11 and the suggested role of AcrF in virulence, pangenome data indicating the absence of *acrF* in ST11 was confirmed at the individual assembly level and with findings subsequently confirmed in the laboratory.

4.2 Aims

- To determine whether the AcrF variant is conserved throughout ST11
- To confirm that the variant of AcrF is non-functional

4.3 Hypothesis

- ST11 encodes a non-functional AcrF protein

4.4 *acrF* is absent from the ST11 pangenome

In Chapter 3 it was shown that the level of conservation of RND and related genes between STs of *E. coli* differs. The genes encoding AcrA and AcrB, which together with TolC form the major RND efflux pump in *E. coli,* were highly conserved across all 20 lineages analysed with each having a mean of 1.2 alleles per ST. Unlike *acrA* and *acrB* which had similar levels of conservation*, acrE* and *acrF* were conserved to different extents across the 20 STs. The greatest number of *acrE* alleles across a lineage was two, this was seen for ST10, ST12, ST69 and ST131. In contrast, *acrF* was more varied with five alleles identified across the ST10 pangenome and four in ST12. The mean number of alleles for *acrF* across the 20 lineages was 1.6 compared to 1.2 for its partner *acrE*. However, zero alleles were identified for *acrF* in ST11 and

when *acrF* was aligned to the assemblies used for the pangenome construction, no genes of 3105 nucleotides, the length of *acrF* in K-12, were identified.

4.5 *acrF* contains a highly conserved insertion which encodes two internal stop codons

To confirm whether *acrF* was indeed absent from ST11 the region downstream of *acrE,* which encodes *acrF* was initially downloaded from 12 randomly selected ST11 genomes available on NCBI. Aligning these sequences against *acrF* from MG1655 revealed an insertion of 45 nucleotides between nucleotides 1860 and 1906 which lengthened the *acrF* gene from 3105 to 3150 nucleotides (Figure 4.1). When translated, this insertion corresponded to 13 additional amino acids plus two internal stop codons (Figure 4.1) suggesting the product would be truncated.

To validate this finding across a larger number of genomes, 2000 assemblies were downloaded from the Enterobase *E. coli/Shigella* database, with 212 removed from the analysis due to the presence of duplicate genomes. The remaining 1788 assemblies were then used to confirm whether the insertion found in 12 ST11 genomes was conserved throughout a wider population. Using the *acrF* sequence from *E. coli* O157:H7 strain TW14359 in the NCBI alignment, it was confirmed that 1745 out of the 1788 assemblies (97.59%) had a gene that matched the length expected for ST11 (3150 nucleotides). Sequence identity of those matches ranged from 99.556% to 100%, with the average percentage identity at 99.94%, showing the high conservation of this insertion within ST11 genomes.



Figure 4.1 *acrF* in *E. coli* ST11 has a 45 nucleotide insertion

a) Alignment of 12 *acrF* sequences downloaded from NCBI database confirms an insertion relative to *E. coli* K-12 substr. MG1655. The *acrF* sequences are homologous up- and downstream of the inserted region b) Translation of these genes (example O157:H7 1) confirms the insertion contains two internal stops codons. Residues in red belong to the insertion.

4.6 AcrF is non-functional in ST11

The *acrF* sequences from *E coli.* K-12 MG1655 and ST11 NCTC 12900 were each cloned into the high-copy vector pET21a following digestion with *Nde*I and *Xho*I. NCTC 12900 is an *E. coli* O157:H7 strain that is Shiga-toxin negative. Sequences were confirmed using Sanger sequencing by Eurofins. An *E. coli* K-12 strain lacking both *acrB* and *acrF* was constructed using homologous recombination (Datsenko and Wanner, 2000) and P1 bacteriophage transduction, with both deletions confirmed again via Sanger sequencing. The *acrF* constructs were then transformed into the *E. coli* K-12 MG1655 Δ*acrB* Δ*acrF* background and the function of each *acrF* construct was determined by measuring susceptibility to a range of dyes and antimicrobials. This assay was done in the absence of *acrB* as the two pumps have very similar substrate profiles and expression of *acrB* can mask the subtle phenotypic changes associated with *acrF* expression (Nishino and Yamaguchi, 2001).

Substrates of AcrF include acriflavine, chloramphenicol, erythromycin, rhodamine 6G and tetracycline (Nishino *et al.*, 2003, Nishino and Yamaguchi, 2001). Deletion of the *acrB* gene increased the susceptibility of the strain to almost all compounds tested (Table 4.1) and further deletion of *acrF* had no additional effect on any of the susceptibilities measured. Due to the selection marker on the pET21a plasmid introduction of the empty vector resulted in a significant increase in the resistance to all β-lactams tested (data not shown). Complementation of the double Δ*acrB* Δ*acrF* knockout with the *acrF* gene from *E. coli* K-12 (EC327) decreased susceptibility to clindamycin hydrochloride from <1 to 8 μg/mL, ethidium bromide from 16 to 128 μg/mL and rhodamine 6G from 16 to 64 μg/mL showing that AcrF from *E. coli* K-12 is expressed and functional. Subtle decreases in susceptibility to several

other compounds including erythromycin and novobiocin were also seen. However, no recovery was seen when the Δ*acrB* Δ*acrF* knockout strain was complemented with the ST11 *acrF* sequence (EC328) that contains the internal stop codons. This confirms the insertion present across ST11 renders AcrF non-functional.

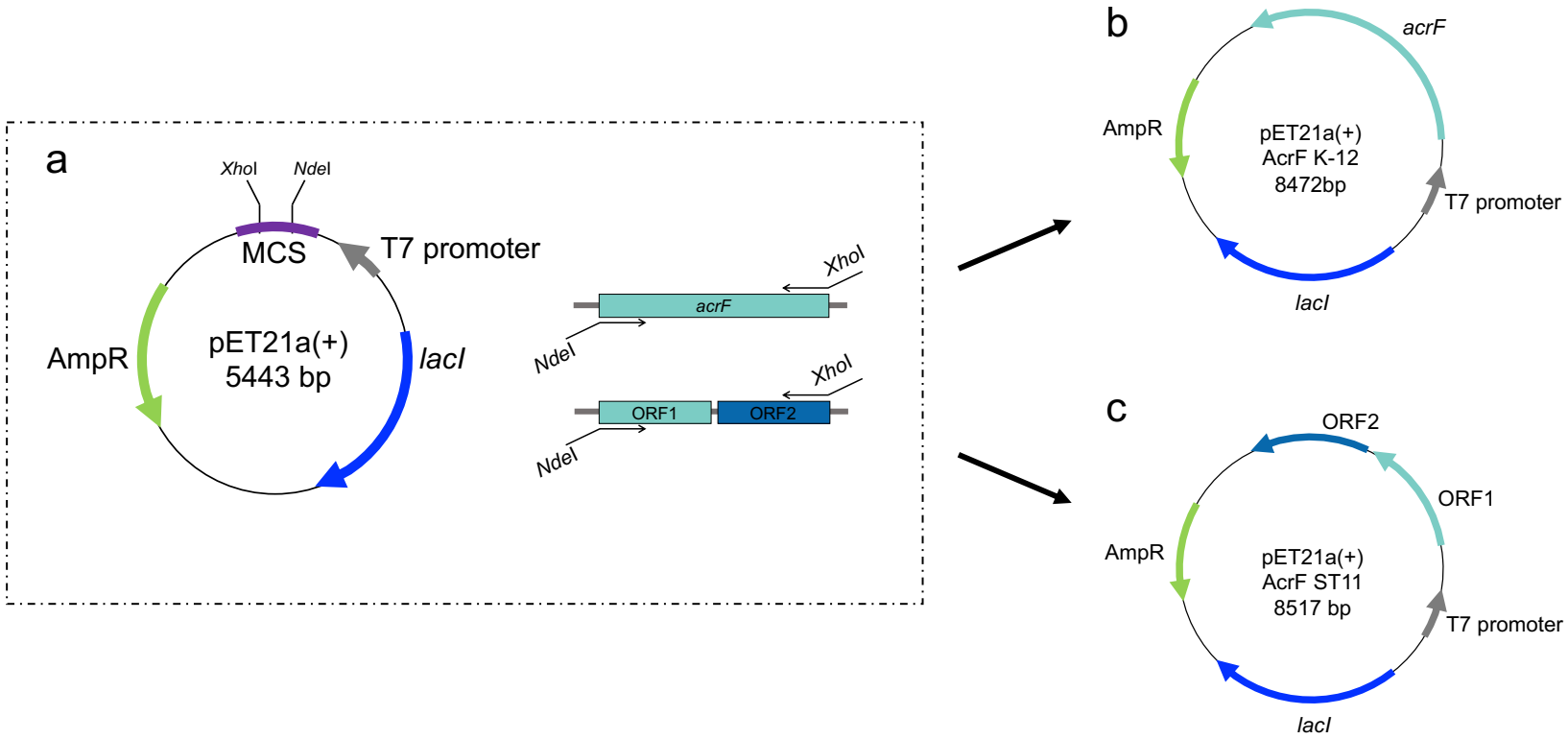


Figure 4.2 Cloning strategy to confirm AcrF in ST11 is non-functional

a) AcrF from *E. coli* K-12 and *E. coli* NCTC 12900 (ST11) were both amplified from the chromosome using primers that incorporated the *Nde*I and *Xho*I restriction sites (top K-12, bottom ST11). The resulting amplimers and pET21a vector were digested with *Nde*I and *Xho*I and ligated together to generate pET21a AcrF K-12 (b) and pET21 AcrF ST11 (c).

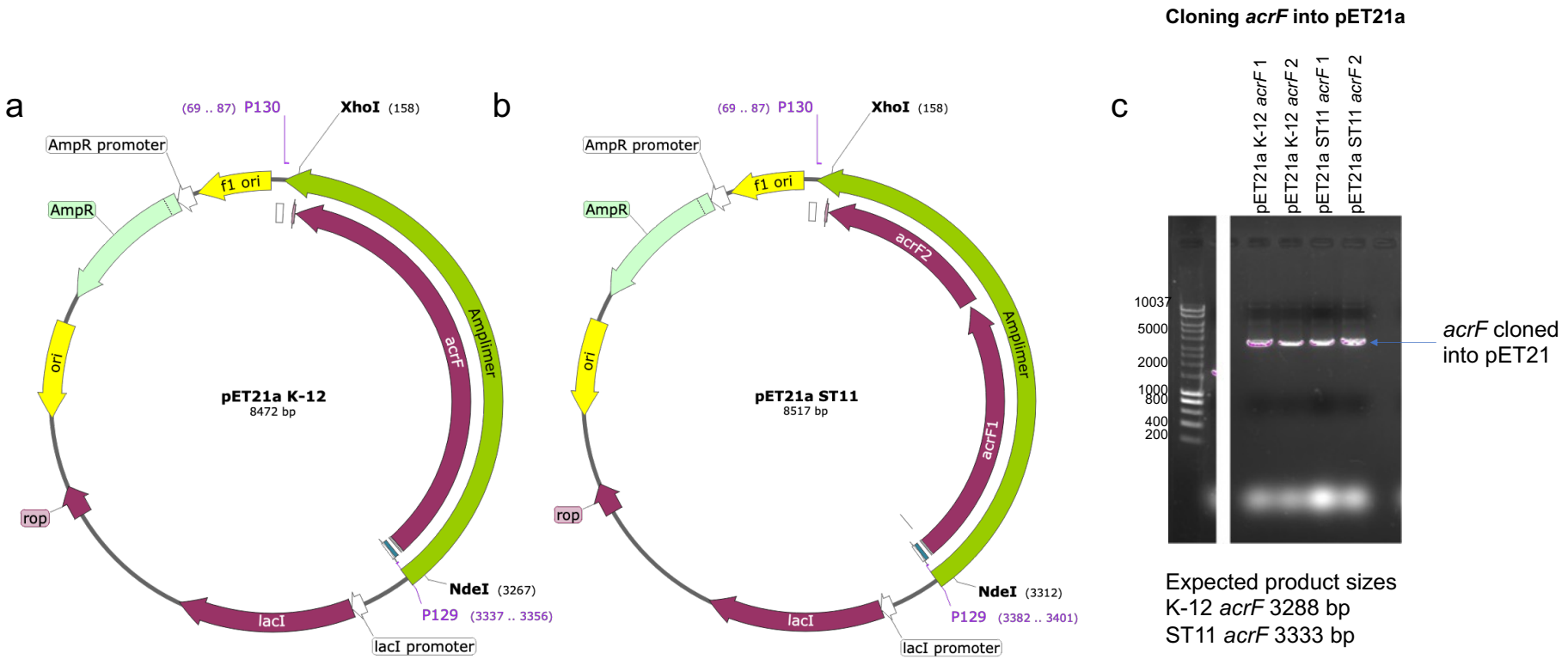


Figure 4.3 Confirmation of *acrF* cloning

The *acrF* gene from *E. coli* K-12 (K-12) and two *acrF* fragments from *E. coli* NCTC 12900 (ST11) were amplified using primers that introduced the *Nde*I and *Xho*I restrictions sites. Following digestion of the *acrF* amplimers, these 3109 and 3154 nucleotide fragments were introduced into the pET21a vector to generate (a) pET21a K-12 and (b) pET21a ST11 constructs. (c) Successful ligation

reactions of pET21a and *acrF* was confirmed using primers that flanked the multiple cloning site, sequences of PCR amplimers were then confirmed with NGS sequencing.

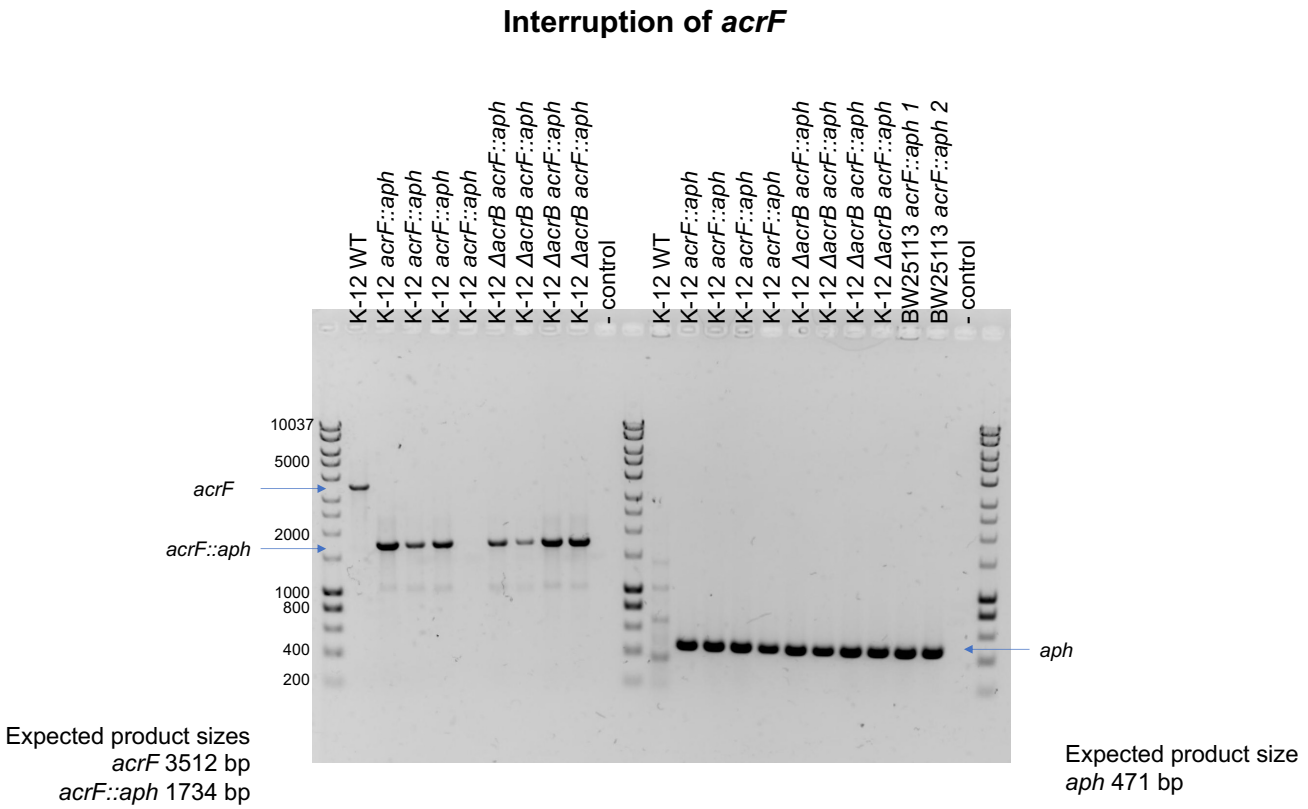


Figure 4.4 Interruption of the *acrF* gene in K-12 and K-12 Δ*acrB*

To confirm the insertion in *acrF* found across ST11 genome assemblies resulted in a non-functional protein, pET21a *acrF* constructs shown in Figure 4.3 were transformed into *E. coli* K-12 Δ*acrF* and *E. coli* K-12 Δ*acrB* Δ*acrF* backgrounds. Construction of Δ*acrF*

strains was achieved using P1 transduction. A phage stock was constructed from the Keio collection *acrF::aph.* The introduction of the *aph* cassette was confirmed via PCR and NGS sequencing. The *aph* resistance cassette was removed using pCP20.

Table 4.1 Susceptibility of K-12 Δ*acrB* Δ*acrF* following complementation with *acrF* from K-12 and ST11

		MIC (µg/mL)												
Identifier	Strain	BAC	CHL	CIP	CLI	ERY	EB	MER	MB	MOX	NAL	NOV	R6G	TET
EC112	ATCC 25922	32	4	0.015	128	64	256	<1	>1024	0.06	4	32	1024	1
EC07	K-12	64	8	0.015	128	64	1024	<1	>1024	0.06	8	128	512	1
EC29	K-12 Δ*acrB*	4	1	<0.008	<1	4	16	<1	4	0.015	2	4	16	0.5
EC118	K-12 Δ*acrB* Δ*acrF*	4	1	<0.008	<1	4	16	<1	32	0.015	2	4	16	0.5
EC233	K-12 Δ*acrB* Δ*acrF* + pET21a	4	1	<0.008	<1	4	16	<1	32	0.015	2	4	16	0.5
EC327	K-12 Δ*acrB* Δ*acrF* + pET21a *acrF* (K-12)	8	1	<0.008	**8**	8	**128**	<1	64	0.015	2	8	**64**	0.5
EC328	K-12 Δ*acrB* Δ*acrF* + pET21a *acrF* (O157:H7)	4	1	<0.008	<1	4	16	<1	32	0.015	2	4	16	0.5

A fold change greater than two is highlighted grey.

BAC benzalkonium chloride, CHL chloramphenicol, CIP ciprofloxacin, CLI clindamycin, ERY erythromycin, EB ethidium bromide, MER meropenem, MB methylene blue, MOX moxifloxacin, NAL nalidixic acid, NOV novobiocin, R6G rhodamine 6G, TET tetracycline.

4.7 Discussion

The aim of this chapter was to confirm whether a functional AcrF is absent from ST11 as suggested in the pangenome analysis.

ST11 is a highly virulent lineage of *E. coli* with an infectious dose as low as 10 cells needed to cause disease due to the secretion of Stx which binds the glycolipid receptor Gb3 (Melton-Celsa, 2014, Nguyen and Sperandio, 2012). Once internalised, Stx induces a ribotoxic stress response leading to apoptosis (Melton-Celsa, 2014). Most infections self-resolve but in severe ST11 cases, patients develop bloody diarrhoea which may progress to HUS and kidney failure (Lim *et al.*, 2010). The main reservoir for ST11 is cattle, where cells replicate asymptomatically at the recto-anal junction due to the absence of the Stx receptor.

Though the exact mechanism by which AcrF contributes to virulence is unknown, inactivation of *acrF* has been shown to reduce cell adherence to human intestinal cells and decrease virulence in a mouse model (Nishino *et al.*, 2006, Blair *et al.*, 2015b). While virulence data is from *Salmonella* species, due to the high amino acid similarity between AcrB in *E. coli* and *Salmonella,* the overlap of substrate profiles and similarities in regulation, it can be hypothesised that they possess similar functions (Wang-Kan *et al.*, 2017, Nishino *et al.*, 2003, Nishino and Yamaguchi, 2001). Moreover virulence has been associated with RND pumps in additional Gram-negative bacteria including *Neisseria gonorrhoeae, Erwinia amylovora, Pseudomonas syringae* and *Vibrio cholerae* (Burse *et al.*, 2004, Jerse *et al.*, 2003, Ichinose *et al.*, 2020, Bina *et al.*, 2018).

While *acrF* in ST11 was not the only gene suggested to be absent across the 20 pangenomes. The only gene that seemed to be a genuinely absent across the STs

was *acrF* in ST11 as in addition to the 0 alleles across the ST11 pangenome, >99% of the 5137 assemblies lacked an AcrF protein sequence. Closer analysis of ST11 assemblies identified an insertion of 45 nucleotides that resulted in a premature termination of translation, therefore truncating the RND protein. The other four genes absent across the pangenomes appeared to be artefacts of the Roary analysis.

A recent review of the evolutionary conservation of efflux pumps across *E. coli* noted that *acrF* was absent from the ST11 core genome and the core genomes of phylogroup C and Shig1 although no investigation of the cause of this was undertaken (Teelucksingh *et al.*, 2020, Abram *et al.*, 2020, Abram *et al.*, 2021). The only RND genes found within the core genome of all phylogroups were *acrB, acrD* and *mdtABC* (Teelucksingh *et al.*, 2020). The findings of Teelucksingh *et al.* (2020) support the data presented in Table 3.1. However, work in this study extends this observation as it shows that the lack of *acrF* in the ST11 pangenome is in fact caused by an insertion.

Due to the well described virulence of ST11 and the known impact of the AcrF homolog in *Salmonella* on infection, it was surprising that ST11 was shown to lack a functional AcrF due to a highly conserved insertion. One explanation for this is due to functional redundancy of RND pumps in Gram-negative bacteria and that the loss of AcrF may not have any significant effect on the bacterial lifecycle. Other RND systems such as AcrAB-TolC or MdtEF-TolC may be upregulated in this lineage to compensate for the lack of AcrEF-TolC function. Indeed several groups have shown AcrF expression increases through mutation in the absence of *acrB* and presence of antimicrobials (Langevin *et al.*, 2020, Zhang *et al.*, 2018, Alon Cudkowicz and Schuldiner, 2019, McNeil *et al.*, 2019). The overlapping of substrate specificity provides a contingency plan should the first line of defence fail (Ghosh and O'Connor,

2017). Furthermore, the main reservoir for this organism is cattle and not humans, so it is plausible that systems linked to extreme acid tolerance or nitrosative conditions such as MdtEF could be more important for a strain that must survive the harsh conditions of the gastrointestinal tract to cause disease.

4.8 Key findings

- The *acrF* gene in the important pathogenic *E. coli* lineage ST11 contains a 45 nucleotide insertion.
- This insertion is conserved across >97% of *E. coli* ST11 assemblies checked.
- Upon translation, the insertion encodes two stop codons that result in protein truncation, rendering the efflux pump non-functional.

4.9 Future work

- Measure the transcription of alternative RND efflux pumps in *E. coli* ST11 using RT-PCR to determine if compensatory overexpression occurs due to the interruption in *acrF.*
- Recreate the 45 nucleotide insertion in an *E. coli* background that possesses a functional AcrF and measure the effect of *acrF* interruption on virulence using the *Galleria mellonella* infection model in the first instance. Using a strain such as *E. coli* ATCC 25922 could help elucidate the role of AcrF in *E. coli* virulence.
- Removal of the 45 nucleotide insertion section and revert the *acrF* gene sequence of ST11 back to that of *E. coli* K-12 and subsequently determine the effect on efflux and antimicrobial susceptibility. Determine whether the insertion sequence results in a fitness cost or has any epistatic effects.

Chapter 5 Characterisation of highly conserved amino acid substitutions present in AcrA, AcrB and TolC

5.1 Background

In Chapter 3 it was shown that across *E. coli,* each component of the AcrAB-TolC efflux pump had a highly conserved polymorphism compared to *E. coli* K-12 MG1655 protein (Figure 3.1 - Figure 3.3). The presence of the alternate amino acids in AcrA, AcrB and TolC were found throughout phylogroups B2, D and F which are traditionally associated with ExPEC infection, while absent from A, B1 and E.

Nonsynonymous single nucleotide changes in RND efflux pumps have previously been shown to confer clinically relevant phenotypic changes e.g. in *S.* Typhimurium, a single substitution of glycine to aspartic acid at position 288 in AcrB results in a clinically relevant phenotype, with increased ciprofloxacin efflux (Blair *et al.*, 2015a). Similarly, a G288S substitution associated with meropenem resistance has been described in a clinical isolate of *A. baumannii* (Hawkey *et al.*, 2018). In the case of *E. coli*, it is possible to predict the AIEC phenotype based upon the presence of SNPs in AIEC and non-AIEC strains (Camprubi-Font *et al.*, 2018).

To investigate the phenotypic impact of AcrA T104A, AcrB H596N, and TolC A233T polymorphisms, these were recreated on pET21a plasmids encoding K-12 AcrA, AcrB and TolC. Complementation with both the K-12 and variant sequences and subsequent phenotypic characterisation allowed direct comparison and investigation of any functional benefit of the polymorphisms.

5.2 Aims

- To investigate the phenotypic impact of the three conserved amino acid substitutions; AcrA T104A, AcrB H596N, TolC A233T.

5.3 Hypotheses

- The AcrA T104A, AcrB H596N and TolC A233T substitutions will confer a phenotypic advantage to *E. coli* causing them to have been selected for across phylogroups B2, D and F
- The introduction of AcrA T104A, AcrB H596N and TolC A233T will alter the drug susceptibility profile of *E. coli* K-12

5.4 The prevalence of specific amino acid substitutions across AcrA, AcrB and TolC in *E. coli* phylogroups

When comparing *E. coli* AcrA, AcrB and TolC protein sequences from 20 ST to the reference sequences from *E. coli* K-12 MG1655, heatmaps (Figure 3.1 - Figure 3.3) demonstrated that substitutions at AcrA 104, AcrB 596 and TolC 233 were highly conserved across proteins that have little overall variation (Table 5.1). The amino acid present at AcrA 104, AcrB 596 and TolC 233 appears to be dependent on the phylogenetic background of the strain. Over the 19601 assemblies included, none of the phylogroup A, B1 or E assemblies contained AcrA T104A, AcrB H596N or TolC A233T substitutions. Yet AcrA T104A was present in 99.997% of phylogroup B2 assemblies and 100% of phylogroup D and F assemblies (Table 5.2). Though present in 80% of phylogroup B2 and 100% of phylogroup D assemblies, the H596N substitution in AcrB was also absent from phylogroup F in addition to A, B1 and E assemblies. The TolC A233T change was present in 99.2% of B2, 41.2% and 48.8%

of phylogroup D and F assemblies respectively (Table 5.2). Interestingly, AcrA T104A, AcrB H596N and TolC A233T co-occur in B2, D and F STs (Figure 5.1). The most prevalent substitution of the three was AcrA T104A as this change was present across all lineages of phylogroup B2, D and F and co-occurred with both AcrB H596N and TolC A233T. However AcrB H596N and TolC A233T did not always co-occur with each other and were only found in the presence of AcrA T104A.

Table 5.1 Prevalence of AcrA T104A, AcrB H596N and TolC A233T across 20 *E. coli* ST

		Prevalence (%)		
Phylogroup	**Sequence Type**	**AcrA T104A**	**AcrB H596N**	**TolC A233T**
A	ST10	0.00	0.00	2.16
A	ST167	0.00	0.00	1.74
B1	ST3	0.00	0.00	0.00
B1	ST17	0.00	0.00	0.00
B1	ST21	0.00	0.00	0.00
E	ST11	0.00	0.00	0.00
B2	ST12	100.00	0.00	98.59
B2	ST14	100.00	100.00	95.16
B2	ST28	100.00	100.00	100.00
B2	ST73	100.00	100.00	99.66
B2	ST95	100.00	0.00	100.00
B2	ST127	100.00	100.00	100.00
B2	ST131	99.97	100.00	99.09
B2	ST141	100.00	100.00	100.00
B2	ST144	100.00	100.00	100.00
B2	ST372	100.00	100.00	100.00
D	ST38	100.00	100.00	6.48
D	ST69	100.00	100.00	75.94
F	ST117	100.00	0.00	97.41
F	ST648	100.00	0.00	0.26

Table 5.2 Prevalence of AcrA T104A, AcrB H596N and TolC A233T across the *E. coli* phylogroups

	Prevalence (%)		
Phylogroup	**AcrA T104A**	**AcrB H596N**	**TolC A233T**
A	0.000	0.000	1.952
B1	0.000	0.000	0.000
E	0.000	0.000	0.000
B2	99.997	80.000	99.249
D	100.000	100.000	41.210
F	100.000	0.000	48.835

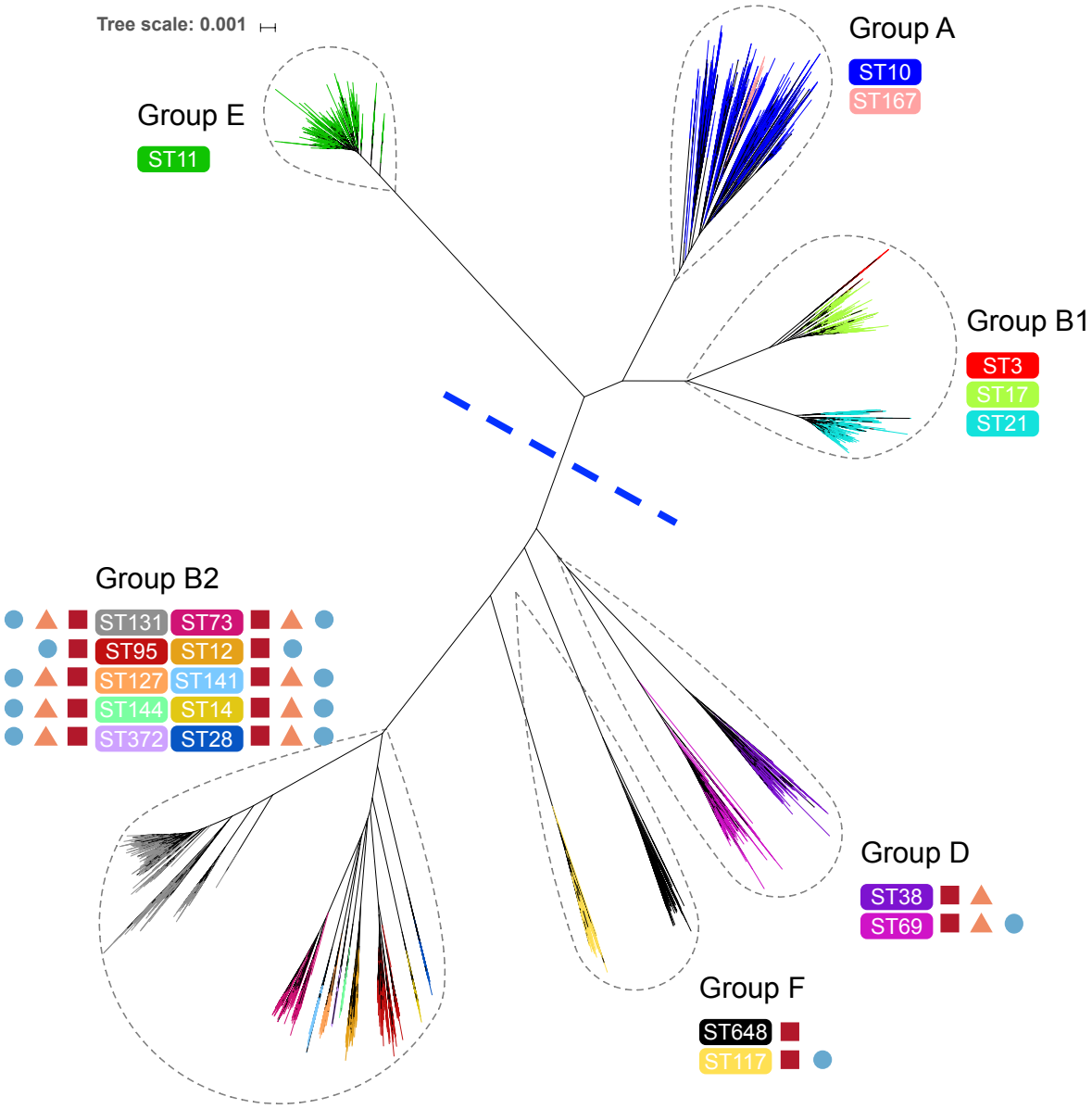


Figure 5.1 Distribution of AcrA T104A, AcrB H596N and TolC A233T across the *E. coli* assemblies used in the pangenome and RND protein conservation analyses

Sequence types in which >70% of assemblies have the substitution are marked with the following symbols; AcrA T104A ■ AcrB H596N ▲ TolC A233T ●.The blue dashed line represents the divide between phylogroups that have the same amino acids as the

reference *E. coli* K-12 MG1655 at residues AcrA 104, AcrB 596 and TolC 233 and those that have the AcrA T104A, AcrB H596N, and TolC A233T substitutions.

5.5 The polymorphisms were not found to alter susceptibility to dyes or antimicrobials.

AcrAB-TolC exports several classes of antimicrobials and in some cases single amino acid substitutions have been shown to have clinically relevant implications (Blair *et al.*, 2015a, Sajib *et al.*, 2021, Yang *et al.*, 2021). The effect of AcrA T104A, AcrB H596N and TolC A233T on susceptibility to a range of AcrAB-TolC substrates was determined in *E. coli* K-12 MG1655.

To do this, pET21a plasmids encoding the *E. coli* K-12 MG1655 *acrA, acrB* and *tolC* sequences and the mutated variants (*acrA* A310G, *acrB* C1785A, *tolC* G697A) were transformed into K-12 lacking either *acrA*, *acrB*, or *tolC,* respectively. Herein the K-12 versions of proteins will be called K-12 AcrA, K-12 AcrB, and K-12 TolC while those containing highly conserved single amino acid variant will be referred to as AcrA T104A, AcrB H596N, and TolC A233T.

AcrAB-TolC is able to confer decreased susceptibility to a wide range of antimicrobials and dyes. The first hypothesis to explain the distribution of the described polymorphisms was that the variants may confer altered susceptibility to clinically useful antimicrobials. To test this the MIC values for *E. coli* K-12 MG1655, strains lacking *acrA, acrB* and *tolC*, and the complemented strains with both K-12 and variant sequences were determined. As previously published, inactivation of either *acrA*, *acrB* or *tolC* increased susceptibility to a range of substrates including novobiocin, erythromycin and chloramphenicol. Complementation of each mutant with either the K-12 versions of AcrA, AcrB, TolC or AcrA T104A, AcrB H596N and TolC A233T was able to restore decreased susceptibility, however no significant differences could be seen between any of the K-12 peptide sequences and the variants (Table 5.3). It is

possible that the MIC methodology is simply not sensitive to detect subtle changes in export of these compounds due to the long incubation period and the doubling dilutions of compounds used.

Table 5.3 Antimicrobial susceptibility of strains complemented with K-12 or variant versions of for *acrA, acrB* and *tolC*

	MIC (µg/mL)																			
Strain	**AZT**	**BAC**	**CAR**	**CEF**	**CHL**	**CLI**	**CV**	**ERY**	**EB**	**FA**	**GEN**	**MER**	**MB**	**MOX**	**NA**	**NOV**	**R6G**	**RIF**	**SPE**	**TET**	**TIC**
ATCC 25922	0.12	32	16	0.06	4	128	32	64	512	1024	0.5	0.03	>1024	0.03	2	32	1024	>32	8	1	8
MG1655	0.06	64	8	0.06	8	128	32	32	1024	1024	0.25	0.03	>1024	0.06	8	128	512	>32	8	1	4
MG1655 Δ*acrA*	0.06	4	2	<0.016	1	2	4	2	16	8	0.12	0.03	8	<0.008	2	4	8	32	4	0.25	1
MG1655 Δ*acrA* + pET21a	0.12	4	>1024	<0.016	1	2	4	2	16	8	0.12	0.03	8	<0.008	2	2	8	32	4	0.25	>1024
MG1655 Δ*acrA* + pET21a::*acrA*	0.12	4	>1024	<0.016	2	4	4	4	64	32	0.12	0.03	16	<0.008	2	4	32	32	4	0.25	>1024
MG1655 Δ*acrA* + pET21a::*acrA* T104A	0.12	8	>1024	<0.016	2	4	4	4	32	32	0.25	0.03	16	<0.008	2	4	32	32	4	0.25	>1024
MG1655 Δ*acrB*	0.06	4	8	<0.016	1	2	4	2	16	8	0.12	0.03	8	<0.008	2	2	8	32	4	0.25	4
MG1655 Δ*acrB* + pET21a	0.12	4	>1024	<0.016	1	2	4	2	8	8	0.12	0.03	8	<0.008	2	2	8	32	4	0.25	>1024
MG1655 Δ*acrB* + pET21a::*acrB*	0.12	16	>1024	0.03	4	32	8	16	128	128	0.12	0.03	256	0.016	2	16	128	32	4	0.5	>1024
MG1655 Δ*acrB* + pET21a::*acrB* H596N	0.12	16	>1024	0.03	4	32	8	16	128	128	0.25	0.03	256	0.016	2	16	128	32	4	0.5	>1024
MG1655 Δ*tolC*	0.25	2	2	0.03	1	1	4	2	4	4	0.12	0.03	8	<0.008	<1	<1	8	32	4	0.25	2
MG1655 Δ*tolC* + pET21a	0.25	2	>1024	0.03	1	1	4	2	4	4	0.12	0.03	8	<0.008	<1	<1	8	32	4	0.25	>1024
MG1655 Δ*tolC* + pET21a::*tolC*	0.25	8	>1024	0.03	2	2	8	16	128	128	0.12	0.03	64	<0.008	<1	16	64	32	4	0.25	>1024
MG1655 Δ*tolC* + pET21a::*tolC* A233T	0.12	8	>1024	0.03	2	2	8	8	128	128	0.12	0.03	64	<0.008	<1	16	64	32	4	0.25	>1024

AZT – aztreonam, BAC – benzalkonium chloride, CAR – carbenicillin, CEF – cefotaxime, CHL – chloramphenicol, CLI – clindamycin, CV- crystal violet, ERY – erythromycin, EB – ethidium bromide, FA – fusidic acid, GEN – gentamicin, MER – meropenem, MB – methylene blue, MOX – moxifloxacin, NA – nalidixic acid, NOV – novobiocin, R6G – rhodamine 6G, RIF – rifampicin, SPE – spectinomycin, TET – tetracycline, TIC – ticarcillin.

5.6 Intracellular accumulation of ciprofloxacin is not affected by the introduction of the conserved residues

Due to the potential insensitivity of the MIC experiments it was hypothesised that a more sensitive approach was needed to measure the effect AcrA T104A, AcrB H596N and TolC A233T on efflux function. Accumulation and subsequent measurement of intracellular levels of fluorescent compounds can be used as indirect method of measuring efflux. Accumulation is indirect because the level at which a compound accumulates within a cell is dependent on a range of factors, such as efflux rate and the rate of dye influx (Blair and Piddock, 2016). Ciprofloxacin was used to measure accumulation instead of ethidium bromide as it is a clinically relevant fluoroquinolone substrate of AcrAB-TolC and its fluorescence can be measured (excitation 280 nm emission 440 nm). With the use of standards of known concentrations, the concentration of ciprofloxacin per cell can be determined from fluorescence levels.

Stains were incubated in the presence of ciprofloxacin for 5, 10 or 20 minutes, cells were then washed and lysed. The supernatant was loaded into a black 96-well microtiter plate and fluorescence measured in a plate reader allowing quantification.

In this study deletion of either *acrA, acrB* or *tolC* had no significant measurable effect on the accumulation of the fluoroquinolone (Figure 5.3). It was therefore not possible to see an effect of complementation with K-12 *acrA,* K-12 *acrB,* K-12 *tolC* or the plasmids encoding AcrA T104A, AcrB H596N and TolC A233T, due to a lack assay sensitivity. As mentioned measurement of accumulation is an indirect method of evaluating efflux function and is less sensitive that measuring active efflux with a fluorescent substrate as accumulation rate depends on many factors such as rate of

efflux. Thus ciprofloxacin accumulation may not be sensitive enough to measure a change.

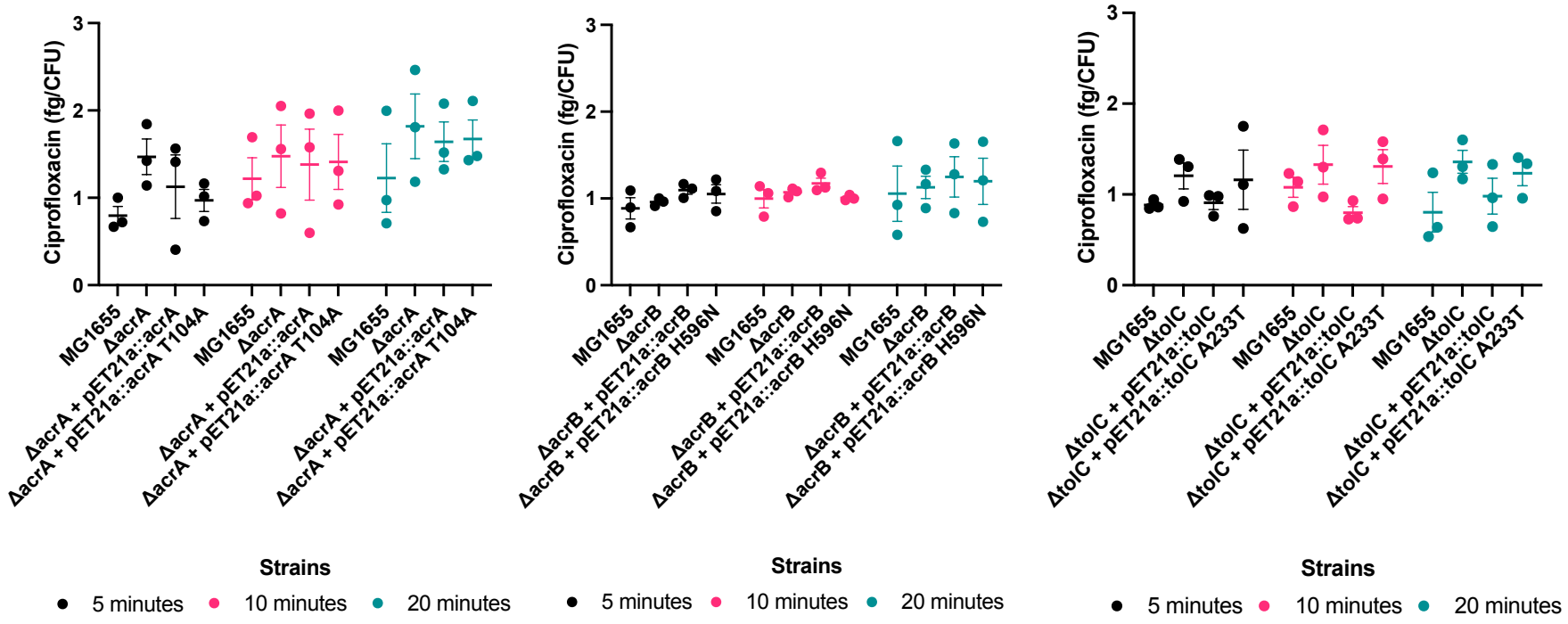


Figure 5.2 Ciprofloxacin accumulation of *E. coli* K-12, Δ*acrA*, Δ*acrB* and Δ*tolC* and strains complemented with both K-12 and amino acid variants

E. coli cultures were incubated in the presence of 10 µg/mL ciprofloxacin for 5, 10 and 20 minutes, the intracellular concentration of ciprofloxacin per CFU was then measured. Scatter plot shows individual biological repeats (*n* = 3), bars show the mean and standard error. No significant differences were measured between strains or conditions (two-way ANOVA, Tukey's multiple comparisons).

5.7 AcrA T104A, AcrB H596N and TolC A233T did not alter the efflux rates of ethidium bromide and 1,2'-dinaphthylamine

In order to measure active efflux directly cells are pre-loaded with a fluorescent dye in the presence of an inhibitor, here CCCP; following re-energisation the rate of efflux is measured by the decrease in fluorescence values as the fluorescent substrate is pumped out, therefore measuring efflux directly. To determine if a subtle phenotypic change was being missed in the accumulation assay, ethidium bromide efflux rate was determined.

In *E. coli*, deletion of *acrA* or *acrB* does not significantly affect the rate of ethidium bromide efflux as both AcrEF and MdtEF can export this dye (Nishino and Yamaguchi, 2001, Wang-Kan, 2018). However, inactivation of *tolC* has a significant phenotypic impact because it is the OMP for five of the six RND pumps in *E. coli.* Therefore, only the rates of efflux for strains lacking *tolC* and those complemented with K-12 TolC and TolC A233T were measured using ethidium bromide. Deletion of *tolC* significantly reduced the rate of efflux of ethidium bromide. Complementation with the pET21a vector encoding K-12 TolC restored function as did the TolC A233T vector (Figure 5.3), as decrease in fluorescence was measured implying ethidium bromide was leaving the cell. However, no significant difference could be seen between the two TolC sequences, K-12 TolC and TolC A233T (Table 5.4).

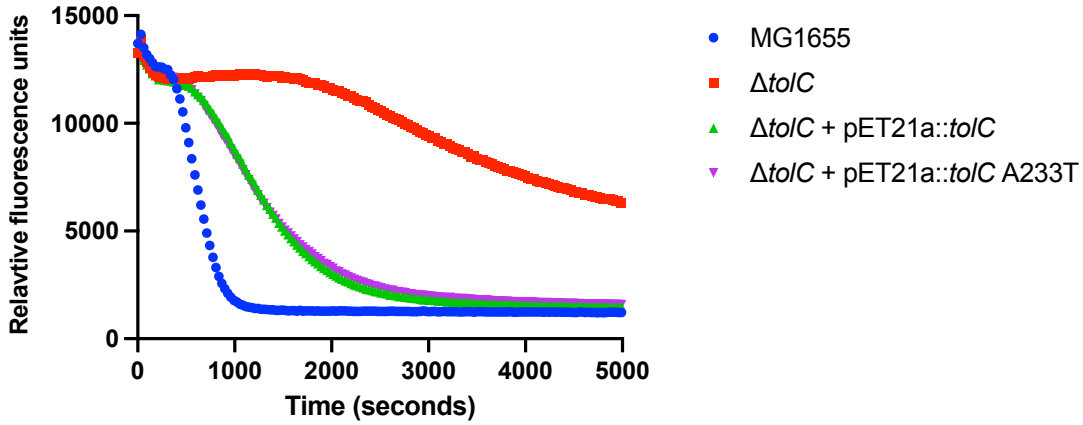


Figure 5.3 Efflux of ethidium bromide in *E. coli* K-12, Δ*tolC* and complemented strains

Efflux kinetics of ethidium bromide for *E. coli* K-12 ●, K-12 Δ*tolC* ■ and strains complemented with the K-12 *tolC* ▲ sequence or A233T variant *tolC* ▼ sequence. Deletion of *tolC* significantly impacted ethidium bromide efflux, however no difference was seen between the efflux of strains complemented with the K-12 or variant *tolC* sequences. Data shown in the mean of three biological replicates. Time taken for fluorescence to decrease by 10%, 25% and 50% were calculated, however no significant difference was detected when comparing both complemented strains (two-way ANOVA, Tukey's multiple comparisons).

Table 5.4 Rate of ethidium bromide efflux in K-12, K-12 *ΔtolC,* K-12 *ΔtolC* pET21a K-12 *tolC* and K-12 *ΔtolC* pET21a *tolC* A233T

	Time (minutes)			
Fluorescence drop	**K-12**	**K-12 Δ*tolC***	**K-12 Δ*tolC* + pET21a::*tolC***	**K-12 Δ*tolC* + pET21a::*tolC* A233T**
10%	2.61 ± 0.41	1.78 ± 0.20	2.76 ± 0.52	1.93 ± 0.01
25%	11.36 ± 3.87	37.73 ± 17.05	17.98 ± 4.57	18.79 ± 6.41
50%	13.71 ± 3.57	90.28 ± 20.55	25.44 ± 5.12	25.65 ± 6.86

Quantification of efflux data presented in Figure 5.3. Data is the mean of three biological replicates ± the standard error of mean. No significant difference was detected when comparing both complemented strains (two-way ANOVA, Tukey's multiple comparisons).

An alternative dye to ethidium bromide that has been used to measure efflux in Gram-negative organisms is the highly sensitive membrane probe 1,2'-dinaphthylamine (1,2'-DNA) (Philippe *et al.*, 2015, Bohnert *et al.*, 2011). Unlike ethidium bromide, 1,2'-DNA is lipophilic and accumulates in the periplasm, because of this it has been suggested that 1,2'-DNA is more relevant for measuring RND efflux as substrates enter RND pumps via the periplasm (Bohnert *et al.*, 2011). Published methods varied greatly meaning a robust method had to be established for use in this project (as described in Methods 2.10.6). Following the establishment of the method, efflux of AcrA, AcrB and TolC complemented strains was measured. Deletion of *acrA, acrB* and *tolC* all prevented compound export. Complementation with either of the *acrA* constructs failed to rescue function and thus no difference could be determined. Both *acrB* sequences had a subtle effect on restoring efflux but like AcrA, neither K-12 AcrB nor AcrB H596N were distinguishable from each other. However, a slight difference in efflux capability was observed between K-12 TolC and TolC A233T suggesting A233T exported the dye 1,2'-DNA at a slightly faster rate (Figure 5.4), however this was not statistically significant (Table 5.5).

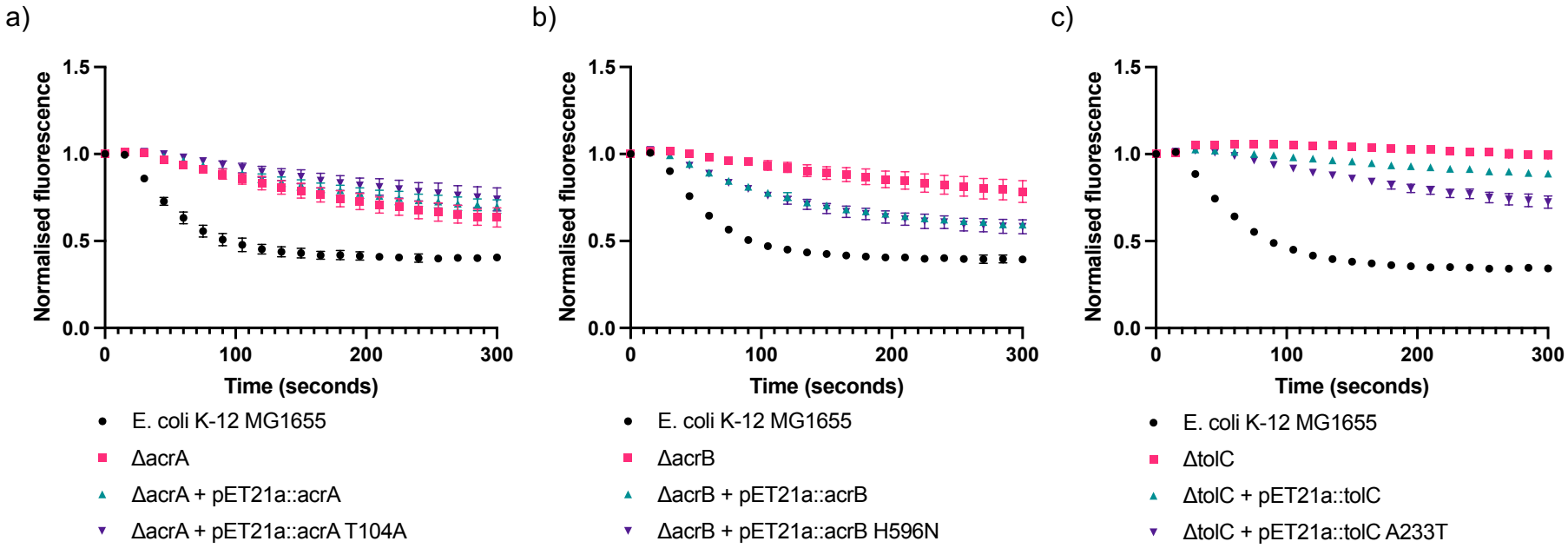


Figure 5.4 Efflux of 1,2'-dinaphthylamine by *E. coli* K-12, Δ*acrA*, Δ*acrB* and Δ*tolC* and complemented strains

Efflux kinetics of 1,2'-DNA for *E. coli* K-12 and *acrA* derivatives (a), *acrB* derivatives (b) and *tolC* (c)

Inactivation of *acrA, acrB* and *tolC* reduced the ability of K-12 to efflux 1,2'-DNA however no significant differences were seen between strains complemented with the K-12 and variants (two-way ANOVA, Tukey's multiple comparisons). Data are the mean of three biological repeats and bars represent the standard error of mean.

Table 5.5 Rate of 1,2'-diphnaphthylamine efflux in *E. coli* K-12 Δ*acrA*, Δ*acrB* and Δ*tolC*, and complemented strains

	Time (seconds)	
Strain	**10% fluorescence drop**	**20% fluorescence drop**
K-12	21.07 ± 5.69	32.86 ± 6.72
K-12 Δ*acrA*	58.61 ± 7.85	121.06 ± 19.92
K-12 Δ*acrA* + pET21a::*acrA*	75.47 ± 10.11	149.41 ± 26.35
K-12 Δ*acrA* + pET21a::*acrA* T014A	111.14 ± 17.08	229.26 ± 70.2
K-12	16.26 ± 5.32	31.34 ± 7.59
K-12 Δ*acrB*	113.71 ± 36.21	257.74 ± 81.35
K-12 Δ*acrB* + pET21a::*acrB*	44.24 ± 3.967	79.19 ± 3.6
K-12 Δ*acrB* + pET21a::*acrB* H596N	35.28 ± 2.03	69.00 ± 6.28
K-12	12.09 ± 0.38	-
K-12 Δ*tolC*	365.37 ± 12.81	-
K-12 Δ*tolC* + pET21a::*tolC*	190.19 ± 36.67	-
K-12 Δ*tolC* + pET21a::*tolC* A233T	76.26 ± 6.77	-

Data are means of three biological repeats ± standard error of mean.

5.8 Curli and cellulose production are not affected by the introduction of AcrA T104A, AcrB H596N or TolC A233T.

The effect of the polymorphisms on other phenotypes that have been previously linked to efflux pumps were also measured. Interruption of RND genes results in the down-regulation of the *csg* operon involved in curli biosynthesis in *S.* Typhimurium (Baugh *et al.*, 2014). Curli are amyloid fibres produced by *E. coli* and *Salmonella* that are involved in attachment and the formation of mature biofilms (Nenninger *et al.*, 2008, Kikuchi *et al.*, 2005). Cellulose is also a biofilm component, with the *bcs* operon highly conserved across the proteobacteria, though "domesticated" *E. coli* K-12 strains have been shown not produce it (Römling and Galperin, 2015, Serra *et al.*, 2013). Congo Red can be used as an indicator of both cellulose and curli production and therefore determine the effects of AcrA T104A, AcrB H596N and TolC A233T on the production of biofilm components.

The cellulose and curli production was assessed for *E. coli* K-12 MG1655, Δ*acrA*, Δ*acrB*, Δ*tolC* and strains complemented with K-12 and variants AcrA T104A, AcrB H596N and TolC A233T. For each strain, 5 µL was plated in triplicate on to LB agar plates supplemented with Congo Red. Following a 48 hour incubation at 30˚C colonies were imaged.

The control strain used, *S.* Typhimurium 14028s produces both curli and cellulose, resulting in a rough dry and red (rdar) phenotype (Römling and Galperin, 2015). The laboratory K-12 strain did not have a wrinkled surface and therefore does not produce cellulose as previously reported (Figure 5.5). Neither the interruption of *acrA, acrB,* and *tolC,* nor complementation with either K-12 or AcrA T104A, AcrB H596N and TolC A233T sequences, had an effect on curli production as all colonies

were smooth and pink. Furthermore, the presence of the pET21a selective agent, ampicillin, had no effect on curli or cellulose production. Therefore, it is unlikely that introduction of the polymorphism would have an effect on biofilm production.

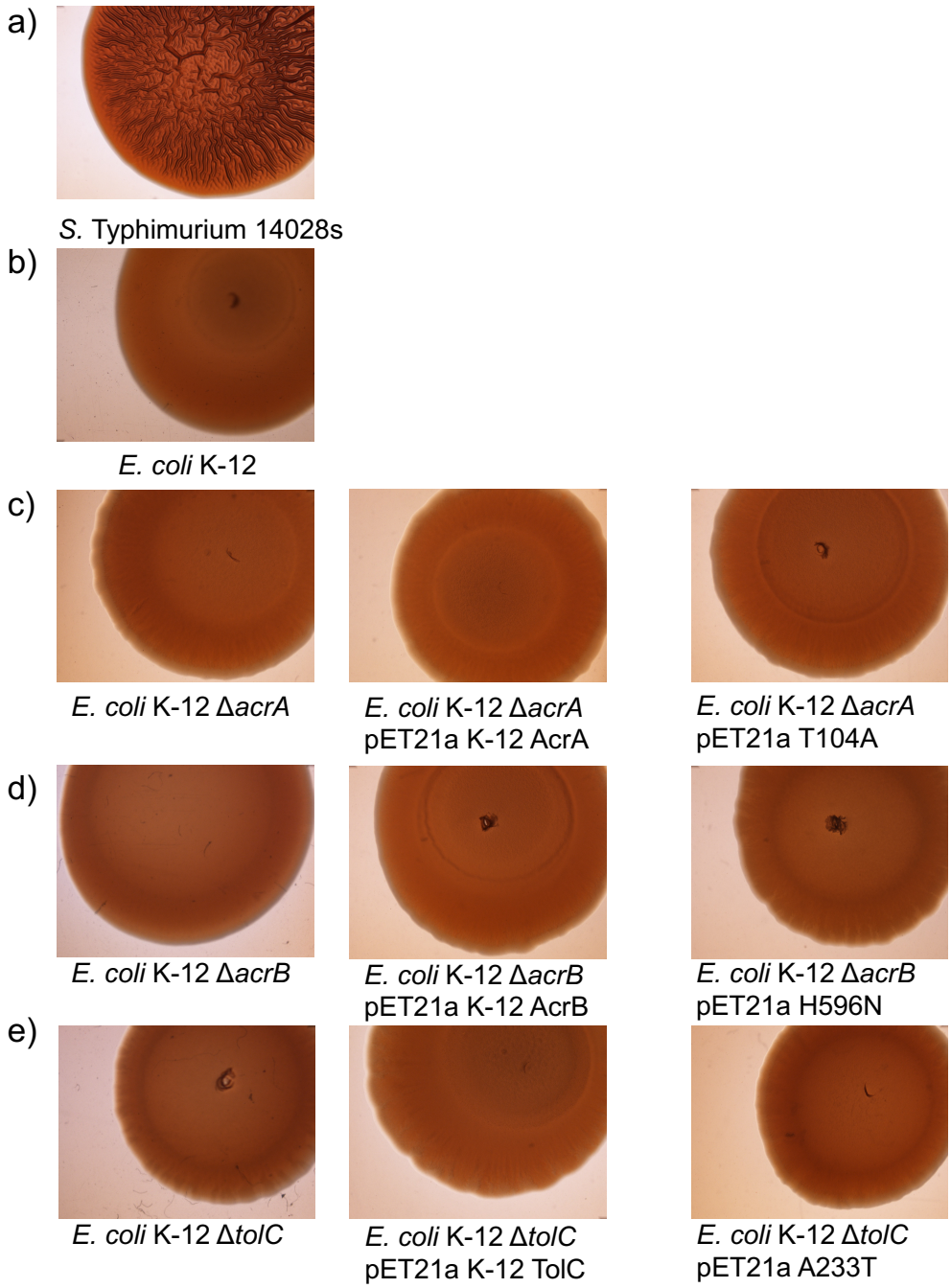


Figure 5.5 Curli and cellulose biogenesis of *E. coli* K-12 MG1655, Δ*acrA*, *ΔacrB*, Δ*tolC* and complemented strains

Cultures were grown for 48 hours at 30˚C on LB agar plates supplemented with 40 µg/mL Congo red. *S.* Typhimurium 14082s was used as a positive control for both curli and cellulose biogenesis (a). *E. coli* K-12 does not produce cellulose and therefore

colonies appear smooth. Interruption of *acrA, acrB* and *tolC,* and complementation with the K-12 and variants sequences (C, D, and E) had no visible effect on curli biogenesis. Images are a representative of three biological repeats.

5.9 Biofilm production is unaffected by the introduction of the conserved mutant residues

Deletion of *acrA, acrB* and *tolC* and complementation with either the K-12 or the AcrA T104A, AcrB H596N and TolC A233T sequences were found to have any effect on colony morphology when grown in the presence of Congo red. However as this is a qualitative assessment of biofilm-associated factors/proteins, the production of biofilm was measured using crystal violet.

Briefly, strains were grown at 30 ˚C with gentle agitation for 48 hours. Bacterial culture was then removed and the biofilms were stained with crystal violet solution. Excess crystal violet was removed and dye staining the biofilm was solubilised using 70% ethanol. The optical density of each well was then measured at 600 nm.

Like in the Congo red assay, inactivation of *acrA, acrB* and *tolC* had no effect on biofilm formation. This could be as biofilm production was low in the K-12 parent strain, subtle changes were not detectable. There was also no change in biofilm production following complementation with pET21a encoding K-12 AcrA, AcrB and TolC or AcrA T104A, AcrB H596N, TolC A233T (Figure 5.6).

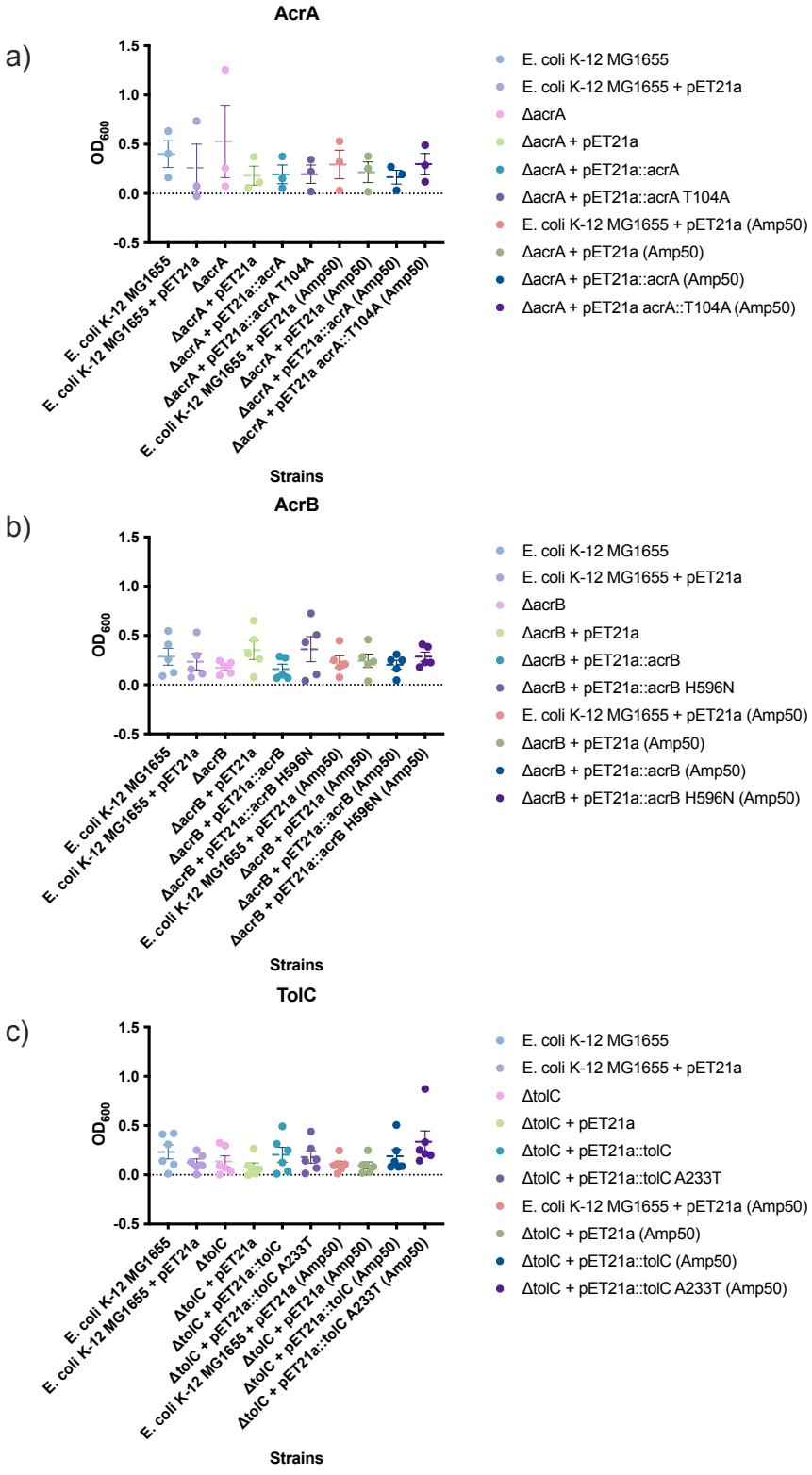


Figure 5.6 Biofilm production in in *E. coli* K-12 MG1655, *ΔacrA, ΔacrB, ΔtolC* and strains complemented with K-12 and amino acid variants

Cultures were grown for 48 hours at 30˚C with gentle agitation, biofilms were stained with crystal violet and the optical density at 600 nm was measured. Interruption of *acrA* (a)*, acrB* (b) and *tolC* (c) had no significant impact of biofilm production, nor did complementation (one-way ANOVA with Tukey's multiple comparisons). Data presented are a minimum of three biological replicates. Error bars show means and standard error of mean.

5.10 Discussion

The aim of the work in this chapter was to identify the phenotypic impact of the AcrA T104A, AcrB H596N and TolC A233T variants to help explain the selection for these substitutions in specific lineages.

5.10.1 Structural location of the variants

As shown in both this Chapter and Chapter 3, *E. coli* K-12 MG1655 has a threonine at position 104 of AcrA which is highly conserved in phylogroups A, B1 and E, yet strains from phylogroups B2, D and F have an alanine. The AcrA T104A substitution has been previously described twice, although its distribution has not been defined and the phenotypic impact of the mutation has not been studied. In the first study the mutation was identified in two isolates obtained from the faecal samples of healthy dairy cows that had been previously treated with 8 μg/mL ceftiofur (Elkins *et al.*, 2010). Though T104A was not characterised the two isolates were serologically typed as O2:H30 and O33:H17 (Elkins *et al.*, 2010). The O2 serotype is commonly associated with ExPEC infections and O2:H30 is an isolate belonging to ST38 and phylogroup D (Dagher *et al.*, 2018, Delannoy *et al.*, 2017). No MLST or phylogroup information could be found for O33:H17. The T104A change was identified also in *E. coli* IAI39, (phylogroup D ST62 serogroup O7:K1) in a study predicting phenotypic trait (however the change was predicted to be neutral) (Galardin *et al.*, 2017). Residue 104 is located at the base of the alpha-helical hairpin of AcrA (Figure 5.7) and has previously been suggested by Marshall (2018) to make direct contact with TolC, however recent experimental data does not support this as AcrA L132, S139 and K140 have been shown to interact with TolC in the tip-to-tip model (Jeong *et al.*, 2016, Xu *et al.*, 2011, Marshall, 2018).

In AcrB, the H596N substitution is present in 8 of the 10 phylogroup B2 STs and both phylogroup D STs in the analysis, while the histidine found in *E. coli* K-12 is maintained in all other STs. These lineages spanned phylogroups A, B1, E and F. This non-synonymous variant has been previously reported in *E. coli* IAI30 which is also phylogroup D though no phenotypic consequence was associated with the H596N substitution (Galardini *et al.*, 2017). Residue 596 is in the PC1 subdomain of AcrB (Figure 5.8), which is involved with substrate uptake (Murakami *et al.*, 2006, Seeger *et al.*, 2006). A histidine to asparagine substitution results in a slight loss of charge at that residue and could therefore affect substrate recognition. However no significant changes were seen between the MIC values for the K-12 AcrB and AcrB H596N although this could be due to a lack of sensitivity of the assay.

The variant present in TolC (A233T) was found in all phylogroup B2 STs. Unlike AcrA T104A and AcrB H596N, this mutation has not been previously reported in the literature and conservation across phylogroups was not as defined. Though it was present in all B2 STs, across the phylogroup F STs 97.41% of ST117 assemblies had the A233T substitution but >99% of ST648 assemblies maintained the K-12 alanine. Only 6.48% of the phylogroup D lineage ST38 had the A233T change and yet 75.94% of ST69 assemblies did. Residue 233 is located in helix 6 of the TolC monomer (Figure 5.9), a region of the protein sits in the periplasm (Koronakis *et al.*, 2000). This residue is not important for channel opening, therefore it is unclear what the structural impact of this change might be (Zgurskaya *et al.*, 2015).

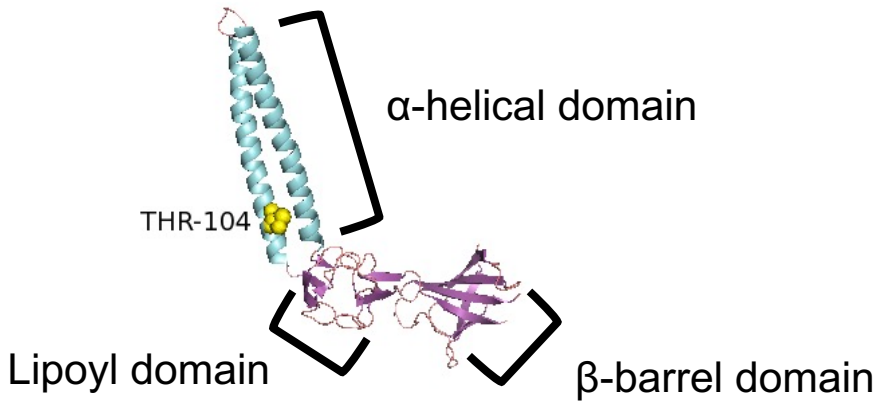


Figure 5.7 Position of T104 within the AcrA monomer

The location of amino acid 104 mapped on an AcrA monomer (PBD: 2F1M).

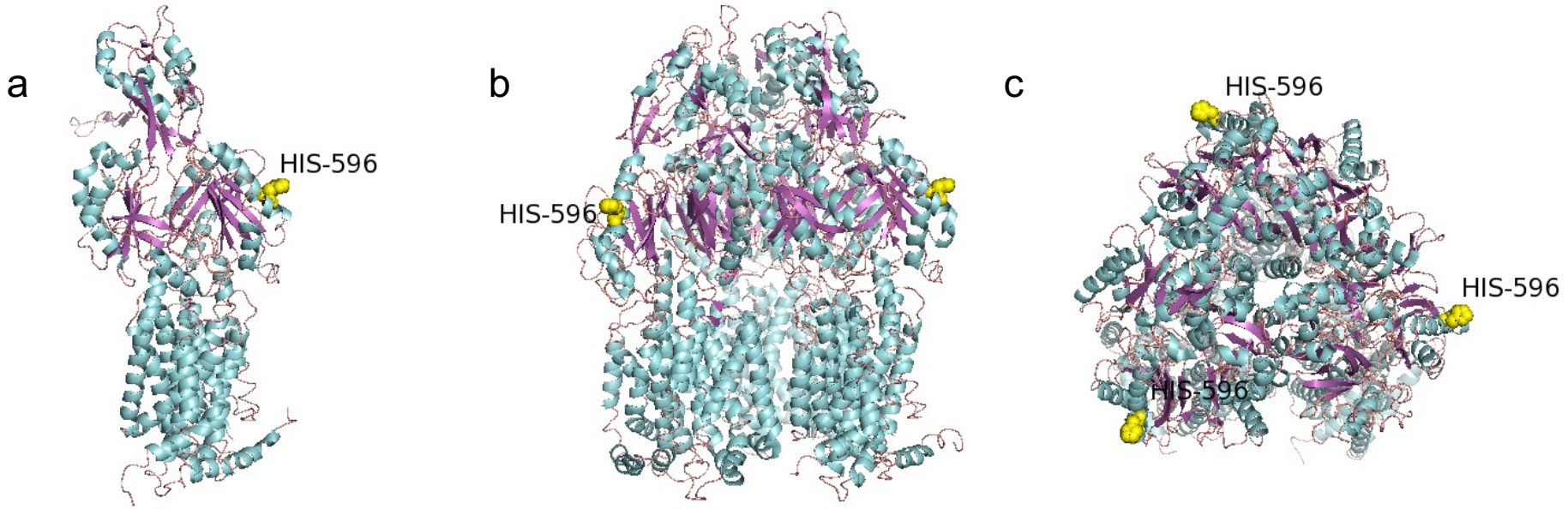


Figure 5.8 Position of H596 within in the AcrB monomer and homotrimer

a) The location of amino acid H596 mapped on an AcrB monomer and homotrimer from the side (b) and above (c) (PBD: 3AOA).

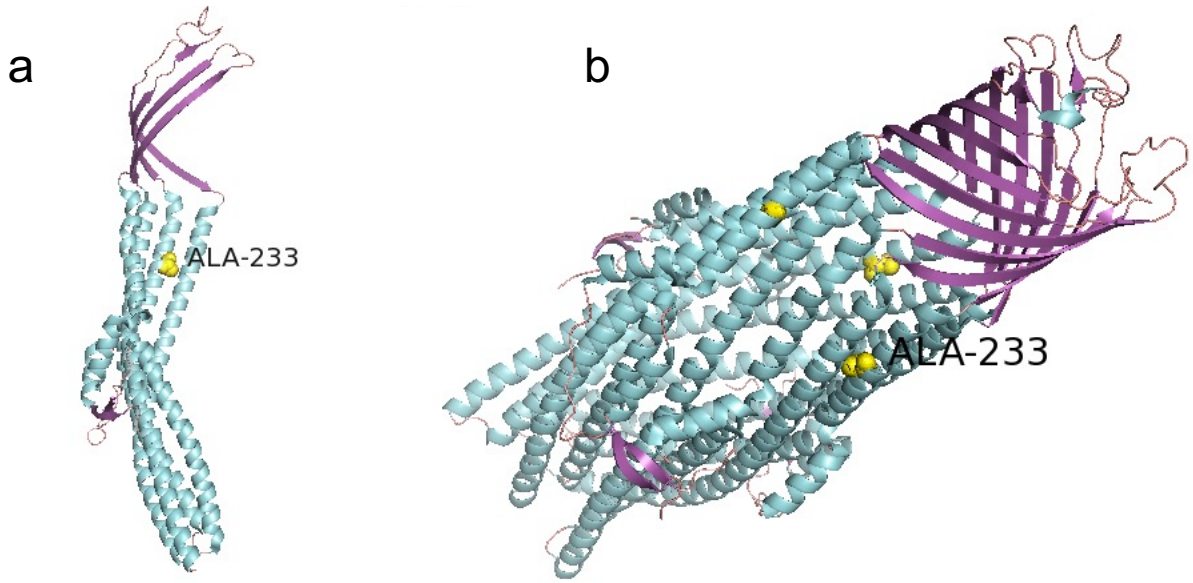


Figure 5.9 Position of A233 within the TolC monomer and trimer

The location of amino acid 233 mapped on an TolC monomer (a) and trimer (b) (PBD: 1EK9)

5.10.2 Phylogeny

Across the *E. coli* population, sequences are clustered into phylogroups based upon genomic similarities. Between phylogroups, genome size has been shown to vary with B1, E and F having the largest genomes and A, B2 and C having the smallest (Horesh *et al.*, 2021). Traditionally, phylogroups A, B1 and C have been associated with more commensal and environmental lifestyles, with B2, D, E and F associated more with pathogenesis (Picard *et al.*, 1999). Due to the high prevalence of the described amino acid polymorphisms in phylogroups B2, D and F, it is possible that these residue changes in AcrA, AcrB and TolC provide a biological advantage that is only required in the environments in which these organisms are found.

As AcrA T104A is the most wide spread variant, perhaps AcrA in combination with another one of the substitutions is needed to see a detectable phenotypic change. Moreover, the background strain for this work could be a problem. This work was done in *E. coli* K-12 MG1655 which belongs to ST10 in phylogroup A, a lineage where none of the variants are found. The genome differences between phylogroups A and those in B2, D, E and F may mean that these substitutions are neutral in this background. Therefore, the amino acid found in K-12 could be introduced to STs that possess the AcrA T104A, AcrB H596N and TolC A233T. Furthermore, in the case of STs that have the AcrA T104A but lack either AcrB H596N or TolC A233T, the missing substitution could be introduced to see what effect, if any, the new polymorphisms have on phenotypes such as efflux.

5.10.3 Phenotype

Considering the location of the AcrB H596N substitution, it was hypothesised that a subtle phenotypic impact may be detected in efflux or MIC experiments due to an impact on substrate recognition. While K-12 and AcrA T104A, AcrB H596N and TolC A233T variants slightly rescued a resistant phenotype, no differences were observed between K-12 and mutant constructs. One potential reason for this is a lack of sensitivity in the assays. MIC experiments in particular lack sensitivity due to the long overnight incubation and due to the doubling dilution, and therefore only large differences in efflux capabilities are be identified with this method. Furthermore, substrates that would have resulted in a phenotypic difference between the K-12 amino acids and the AcrA T104A, AcrB H596N and TolC A233T sequences may have simply been missed hence no changes were detected in this work. A lack of sensitivity could also explain why no phenotypic differences were identified during efflux assay experiments, the phenotypic difference caused by the single amino acid changes may be too subtle for the method. Additionally, only two substrates were used for efflux assays, a more appropriate substrate may be available.

Another possibility is that these mutations may have no phenotypic effect on the function of the pump but instead impact protein complex stability. If this is the case, then measuring the rate of active efflux or a strain's ability to produce curli and biofilms will not demonstrate the advantageous nature of the amino acid.

In summary, an interesting phylogenetic distribution of these mutations has been described but at present, it has not been possible to determine why these have been selected for. Further work, with more sensitive assays may be needed to elucidate the impact of these mutations.

5.11 Key findings

- The sequences of AcrA, AcrB and TolC from *E. coli* K-12 is not representative of all *E. coli*
- AcrA T104A, AcrB H596N and TolC A233T are common amino acid substitutions across *E. coli* when using K-12 as a reference.
- These substitutions are found throughout phylogroups B2, D and F.
- No phenotypic impact of the polymorphisms has been detected.

5.12 Future work

The amount of laboratory work in this chapter was curtailed by restrictions due to the COVID pandemic. With more time the following experiments would have been done.

- Introduce combinations of the variants into an *E. coli* K-12 MG1655 background. As AcrA T104A is the most dominant of variants it would be interesting to see what affect if any AcrA T104A/AcrB H596N, AcrA T104A/TolC A233T, and AcrA T104A/AcrB H596N/TolC A233T have on K-12.
- In addition to gain of function experiments such as those above, reversion mutations could be made in a strain which has all three variants. Here the opposite changes would be made, AcrA A104T, AcrB N596H and TolC T233A, in a background such as the MIC reference strain *E. coli* ATCC 25922 which belongs ST73 and phylogroup B2 which already possesses AcrA T104A, AcrB H596N and TolC A233T.
- Phylogroups B2, D and F are typically more associated with ExPEC infection therefore this could be linked to the prevalence of AcrA T104A, AcrB H596N

and TolC A233T. The effects of K-12 gaining the AcrA T104A, AcrB H596N and TolC A233T substitutions on virulence could be studied using the *Galleria mellonella* model. Alternatively, AcrA T104A, AcrB H596N and TolC A233T could be reverted to the K-12 sequences in a strain such as ATCC 25922 that is highly virulent in *G. mellonella* and assess if the K-12 amino acids reduced the virulence of strains.

Chapter 6 Identification, conservation and characterisation of the novel *eefRABCD* operon across *E. coli*

6.1 Background

The current dogma is that *E. coli* have 6 RND efflux systems regardless of phylogroup or ST. However, this is largely based upon work done to characterise RND pumps in *E. coli* K-12 or K-12 derivatives (Nishino and Yamaguchi, 2001, Du *et al.*, 2014, Wang *et al.*, 2017). In Chapter 4 the notion that all *E. coli* have six RND pumps was demonstrated to be incorrect as the important human pathogenic lineage ST11 was shown to have a conserved insertion in *acrF* resulting in a truncated and non-functional protein.

Previous work by the group identified an additional TetR regulator, EefR, present in 4 of the 10 *E. coli* proteomes included in the analysis (Colclough *et al.*, 2019). The *eefR* gene was found upstream of *eefABC* which encodes an RND system first described in *Enterobacter aerogenes,* now *Klebsiella aerogenes.* The *eef* operon has only been described once in *E. coli*. This strain, *E. coli* SMS-3-5, was isolated from heavily polluted water in North America and was found to have a five gene *eef* operon encoding the *eefR* regulator followed by *eefABCD* encoding a PAP, RND pump, OMP and MFS pump (Fricke *et al.*, 2008). However Fricke *et al.* (2008) did not characterise the *eef* operon.

6.2 Aims

- To determine the prevalence and conservation of *eefRABCD* across *E. coli.*

- To investigate the function of EefABC.

6.3 Hypotheses

- The *eefRABCD* operon will be present across selected *E. coli* lineages.

6.4 The EefABC RND efflux pump in *E. coli* is not limited to *E. coli* SMS-3-5

Previous work that identified the *eefRABCD* operon in several *E. coli* strains had a very limited sample size of 10 *E. coli* proteomes (Colclough *et al.*, 2019). Therefore, the initial aim of this study was to determine the prevalence of the *eef* operon across the diversity of *E. coli*. In the first instance, the sequence of *eefB* (the RND pump component of the system) from *E. coli* SMS-3-5 was used to search across *E. coli* genomes (Taxid 562) available on NCBI. SMS-3-5 was selected as this was one of the strains originally identified by Colclough *et al.* (2019). Furthermore, this is the only strain in which *eefABC* has been reported elsewhere (Fricke *et al.*, 2008). This quick search suggested that the *eefB* was present across a range of well characterised *E. coli* strains including CFT073, O157:H7 and ATCC 25922. As a result, it was hypothesised that the *eef* operon would also be found in a diverse range of *E. coli* strains. However, *eefB* was not present in *E. coli* K-12 MG1655 which is commonly used in RND studies and may explain why EefABC has not been previously characterised.

6.5 The *eefRABCD* operon is highly prevalent across *E. coli*

The sequences of *eefR, eefA, eefB, eefC* and *eefD* were downloaded from *E. coli* ATCC 25922 (CP009072.1), these were annotated in the genome as locus tags DR76_3503 through to DR76_3499 respectively (Table 6.1). A custom BLAST

database was created using these five gene sequences allowing the presence and conservation of the *eef* genes across the downloaded *E. coli* assemblies to be determined using Abricate (v. 0.9.8). *E. coli* assemblies from 38 ST and eight phylogroups were downloaded from Enterobase (Alikhan *et al.*, 2018). Of the 20,013 downloaded assemblies, 766 were removed from the analysis due to being duplicates (Table 6.2). The presence and coverage of the *eef* genes across *E. coli* was determined using Abricate. As assemblies were used instead of fully constructed genomes, genes were occasionally split over contigs which are sequences generated from multiple overlapping sequence fragments. In cases where genes were split across assembly contigs, the assembly was removed from the analysis as it could not be confirmed whether the genes were split over neighbouring contigs and an artefact of sequencing or gene duplications. This resulted in 111 assemblies being removed from the Abricate dataset, resulting in a final 19,136 assemblies (Table 6.2).

Presence of the *eef* operon was found to be confined to specific phylogroups. It was present in all phylogroups B2, F and G assemblies, and in the majority of phylogroup D and E assemblies. In all lineages of *E. coli* that encode the *eef* operon the gene coverage – which relates to how much of a query is covered by the target sequence and gives a relative length – was found to be extremely high (Table 6.3). This suggested an all or nothing conservation because in assemblies that contained the operon, genes were found to have >98% gene coverage to the reference operon from ATCC 25922. Gene identity refers to how many characters are identical between two sequences. Average percentage identity across all genes was above 97% with the only exception being *eefC* in ST354, which had a 96.26% identity to the *eefC* reference

used (Table 6.4). However, the *eefRABCD* operon was completely absent from phylogroups A, B1 and C1.

To determine the evolutionary relationship between those *E. coli* ST that encode the operon and those that do not, five assemblies from each ST were selected at random to construct a phylogenetic tree. To root the tree 27 *Escherichia fergusonii* assemblies were downloaded from Enterobase and duplicate sequences were removed. From the resulting 22 *E. fergusonii* assemblies, five were selected randomly. These *E. coli* and *E. fergusonii* genomes were annotated using autoprokka.py before pangenome construction using Roary. The subsequent core_gene_alignment.aln file for the *E. coli* and *E. fergusonii* assemblies was used to generate a phylogenetic tree using RAxML.

The resulting tree (Figure 6.1) demonstrated a clear phylogenetic split between phylogroups that encode the *eef* operon and those that do not. *E. coli* phylogroups A, B1 and C are conventionally thought of as commensal and here the *eef* operon was completely absent. The operon was present across phylogroups B2, D, E, F and G. Phylogroups B2 and D possess a greatest number of virulence determinants, have been shown to have more defined pathovars and are most commonly isolated from ExPEC infections (Escobar-Páramo *et al.*, 2004, Picard *et al.*, 1999).

Table 6.1 The *eefRABCD* operon and its proposed products

Hypothesised gene	Location in CP009072.1 (complement strand)	Product in CP009072.1 (NCBI exact annotation)	Locus tag in CP009072.1	Gene length (nucleotides)
eefR	3742500..3743066	bacterial regulatory, tetR family protein	DR76_3503	567
eefA	3741230..3742351	efflux transporter, RND family, MFP subunit	DR76_350	1122
eefB	3738123..3741230	RND transporter, hydrophobe/amphiphile efflux-1 family protein	DR76_3501	3108
eefC	3736746..3738119	efflux transporter, outer membrane factor (OMF) lipo, NodT family protein	DR76_3500	1374
eefD	3735574..3736737	drug resistance transporter, Bcr/CflA subfamily protein	DR76_3499	1164

Table 6.2 *E. coli* assemblies used in this analysis

Phylogroup	ST	Downloaded assemblies	Duplicates removed	Split operons removed	Final assemblies
A	ST10	1758	0	1	1757
	ST119	110	0	0	110
	ST167	115	0	0	115
	ST540	136	14	0	122
	ST617	181	0	0	181
B1	ST17	1884	0	0	1884
	ST21	2411	0	0	2411
	ST53	5	0	0	5
	ST84	2	0	0	2
	ST156	210	10	0	200
	ST328	54	0	0	54
	ST453	142	6	0	136
	ST655	856	238	1	617
	ST678	116	9	0	107
B2	ST12	479	41	3	435
	ST14	46	0	0	46
	ST73	916	0	7	909
	ST95	751	0	0	751
	ST127	231	0	1	230
	ST131	3186	0	30	3156
	ST144	85	0	0	85
C	ST23	384	19	0	365
	ST88	454	33	0	421
	ST90	124	8	0	116
	ST410	596	23	2	571
D	ST38	815	29	10	776
	ST69	696	0	0	696
	ST405	525	35	9	481
	ST963	55	2	0	53
E	ST11	1490	132	30	1328
	ST182	76	1	3	72
	ST350	34	4	1	29
	ST1266	15	0	2	13
F	ST59	175	18	1	156
	ST62	217	122	2	93
	ST354	266	22	7	237
	ST648	379	0	1	378
G	ST738	38	0	0	38

Table 6.3 Presence and subsequent gene coverage of *eefRABCD* across the *E. coli* assemblies with reference to the *E. coli* ATCC 25922 operon

Phylogroup	ST group	Assemblies (*n*)	Mean sequence coverage to ATCC 25922 *eef* operon (%)				
			eefR	*eefA*	*eefB*	*eefC*	*eefD*
A	10	1757	0.06	0.06	0.06	0.06	0.00
	119	110	0.00	0.00	0.00	0.00	0.00
	167	115	0.00	0.00	0.00	0.00	0.00
	540	122	0.00	0.00	0.00	0.00	0.00
	617	181	0.55	0.55	0.55	0.55	0.55
B1	17	1884	0.00	0.00	0.00	0.00	0.00
	21	2411	0.00	0.00	0.00	0.00	0.00
	53	5	0.00	0.00	0.00	0.00	0.00
	84	2	0.00	0.00	0.00	0.00	0.00
	156	200	0.50	0.00	0.00	0.00	0.00
	328	54	0.00	0.00	0.00	0.00	0.00
	453	136	0.00	0.00	0.00	0.00	0.00
	655	617	0.00	0.00	0.00	0.00	0.00
	678	107	0.00	0.00	0.00	0.00	0.00
B2	12	435	100.00	99.84	99.77	99.77	100.00
	14	46	100.00	100.00	100.00	100.00	100.00
	73	909	100.00	100.00	100.00	100.00	99.97
	95	751	99.99	100.00	100.00	100.00	100.00
	127	230	99.98	100.00	100.00	100.00	100.00
	131	3156	99.87	99.87	99.91	99.91	99.95
	144	85	99.99	100.00	100.00	100.00	100.00
C	23	365	0.27	0.27	0.08	0.00	0.00
	88	421	0.00	0.00	0.02	0.00	0.00
	90	116	0.00	0.00	0.00	0.00	0.00
	410	571	0.00	0.00	0.00	0.00	0.00
D	38	776	99.69	99.74	99.74	99.74	99.74
	69	696	0.00	0.00	0.00	0.00	0.00
	405	481	99.97	100.00	100.00	100.00	100.00
	963	53	99.99	100.00	100.00	100.00	100.00
E	11	1328	99.99	100.00	99.90	100.00	100.00
	182	72	0.00	0.00	0.00	0.00	0.00
	350	29	99.68	100.00	100.00	100.00	100.00
	1266	13	100.00	100.00	100.00	100.00	100.00
F	59	156	99.32	99.36	99.36	99.21	99.36
	62	93	98.92	98.91	98.92	98.92	98.92
	354	237	98.50	99.07	99.16	99.16	99.35
	648	378	100.00	100.00	100.00	100.00	100.00
G	738	38	100.00	100.00	100.00	100.00	100.00

Cells highlighted red indicate the presence of the *eef* gene across that *E. coli* ST.

Table 6.4 Conservation as percentage nucleotide identity of *eefRABCD* with reference to *E. coli* ATCC 25922 across phylogroups of *E. coli* found to possess the operon

Phylogroup	ST group	Assemblies (*n*)	Mean nucleotide identity to *eefRABCD* in ATCC 25922 (%)				
			eefR	*eefA*	*eefB*	*eefC*	*eefD*
B2	12	435	99.82	99.98	99.45	99.25	99.40
	14	46	98.91	98.49	99.41	99.12	99.57
	73	909	100.00	100.00	100.00	100.00	99.97
	95	751	99.82	99.91	99.77	99.45	99.47
	127	230	97.39	97.90	97.88	97.69	97.75
	131	3156	99.58	99.45	99.52	99.07	99.48
	144	85	99.65	99.91	99.74	99.70	99.74
D	38	776	98.86	98.40	98.16	97.13	98.71
	69	696	-	-	-	-	-
	405	481	99.45	98.32	99.00	99.20	99.14
	963	53	99.12	98.66	98.42	97.38	98.97
E	11	1328	99.12	97.95	98.06	97.53	98.37
	182	72	-	-	-	-	-
	350	29	99.47	98.03	98.38	97.62	98.25
	1266	13	99.47	98.22	98.46	97.58	98.19
F	59	156	98.30	98.51	98.46	98.39	98.84
	62	93	98.05	97.68	98.32	97.70	98.16
	354	237	97.68	98.01	98.68	96.26	99.24
	648	378	99.64	99.47	99.03	98.91	99.47
G	738	38	99.65	98.84	98.91	99.34	99.83

The extent of *eef* gene conservation as percentage nucleotide identity across the phylogroups B2, D, E, F and G are highlighted as follows; ≥ 99% ■, ≥ 98% ■ , ≥ 97% ■ , ≥ 96% ■. Where *eef* was not present, cells contain "-".

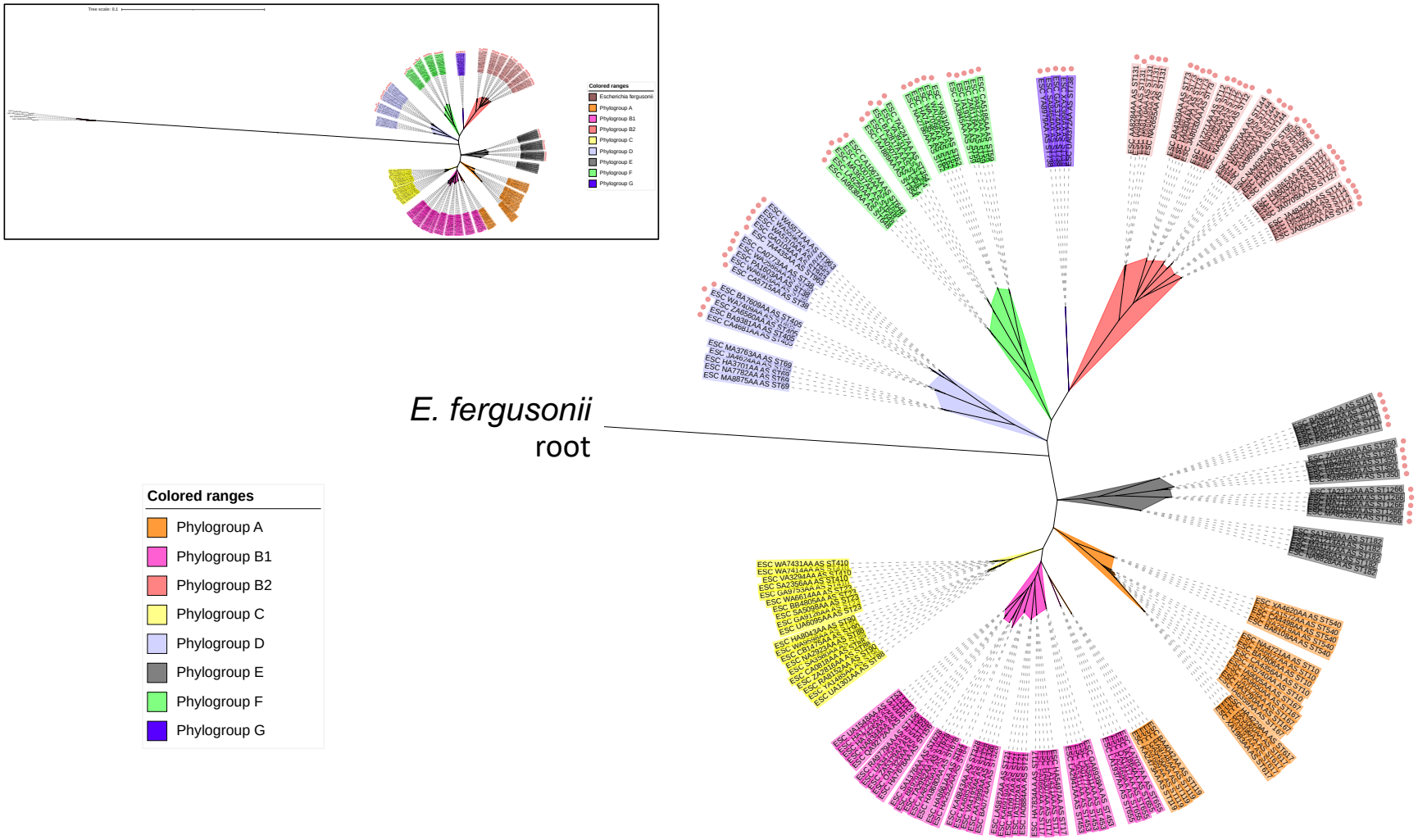


Figure 6.1 Phylogenetic structure of the assemblies used in this analysis and the distribution of *eefRABCD*

Tree created with five assemblies per ST (except ST84), using RAxML. *E. coli* assemblies were rooted with *Escherichia fergusonii,* the full tree including this root is shown in the insert at the top left-hand side, the main image contains the same tree with the root cropped. Leaves are annotated with the ST to which the assembly belongs and coloured based on phylogroup. Colours used to highlight phylogroups A, B1, B2, C, D, E, F and G are shown in the legend to the left of the tree. STs found to possess the *eef* operon possess ● at the end of the leaf.

6.6 Despite belonging to phylogroups D and E, ST69 and ST182 do not possess the *eef* operon

In both phylogroups D and E a single ST was found to not encode the *eefRABCD* operon; these were ST69 and ST182, respectively. ST69 is a clonal complex UPEC lineage that causes persistent UTI infections (Woodford *et al.*, 2011, Yamaji *et al.*, 2018). Furthermore, this lineage of *E. coli* often accounts for a large proportion of community acquired drug resistant infections (Yamaji *et al.*, 2018, de Souza da-Silva *et al.*, 2020, Nüesch-Inderbinen *et al.*, 2017). Despite this and the fact that ST69 clearly clusters with phylogroup D, across the 696 assemblies ST69 was not found to possess any of the *eef* genes.

ST182 is an ETEC lineage commonly associated with foodborne outbreaks (Nishikawa *et al.*, 1998, Nishikawa *et al.*, 1995, Cho *et al.*, 2014). ST182 isolated from airline waste has been demonstrated to be a reservoir for AMR-related genes, carrying ESBLs and mutations in *gyrA* resulting in β-lactam and fluoroquinolone resistance respectively (Jarocki *et al.*, 2021). Yet despite belonging to phylogroup E, none of the 72 ST182 assemblies possess the *eef* operon.

6.7 *eefRABCD* is conserved across phylogroups B2, D, E, F and G but has been lost in A, B1 and C

Abricate and phylogenetic data implied that the presence or absence of *eefRABCD* was an evolutionary trait, one that could have potentially occurred at the diversification of *E. coli,* i.e. when different lineages specialised. The genomic context of the operon was therefore determined to try and ascertain if the operon had been gained independently by lineages and STs, or whether the operon had been lost.

Reference sequences from NCBI RefSeq database were downloaded from 7 of the 38 STs due to the limitations in available reference sequences for the lineages included in this analysis. The seven assemblies comprised of four assemblies which possessed the operon and three which did not. Genomes were aligned and visualised using EasyFig. In each ST the *eef* operon was found between the essential gene *fabI* and the non-essential gene *sapF* (Figure 6.2) (Goodall *et al.*, 2018). FabI is an enoyl-[acyl-carrier-protein] reductase that is involved in fatty acid production, while SapF is a putrescine export protein belonging to the *sapBCDF* system (Sugiyama *et al.*, 2016, Heath and Rock, 1995).

In strains lacking the *eef* operon, the upstream and downstream gene sequences remained highly similar to those that possessed the efflux pump. However, in place of the 7493 nucleotide *eefRABCD* operon, a single open reading frame (ORF) was annotated as *ycjD* (Figure 6.2). The *ycjD* ORF was 354 nucleotides long and predicted to run in the opposite direction to the *eef* operon. The *fabI* through to *sapF* genome region from assemblies lacking the *eef* operon was aligned to those found to possess the operon, this indicated the intergenic region in strains lacking *eef* contained conserved remnants highly homologous to the *eef* operon (Figure 6.3). This suggests that ancestral strains encoded the *eef* operon but it has been degraded or lost from strains that now do not have it. Interestingly, the same remnants of *eef* genes were conserved across phylogroups A, B1 and C which lack the operon suggesting that loss has occurred through gene degradation as opposed to *eef* being gained in phylogroups B2, D, E, F and G. The remnants of *eef* sequences conserved in the ST69 and ST182 assemblies (from phylogroups D and E respectively) were also highly similar to those in phylogroups A, B1 and C, though in both ST69 and ST182 additional fragments of

the *eef* operon have also been lost. This is interesting as all other phylogroup D and E assemblies included in the analysis possessed the pump. Based upon their position in the phylogenetic tree and the different fragments of retained remnants of the *eef* operon it is likely that these two lineages have lost the operon in an additional events separate from A, B1 and C lineages.

Due to the presence of specific and highly conserved remnants of *eef* genes in strains that lack the operon, it was hypothesised that the region was under a strong selection pressure. One hypothesis for this high conservation is the proximity of *eef* to the location of the *fabI* promoter which is an essential gene in K-12 and catalyses a crucial regulatory step in the biosynthesis of fatty acids (Goodall *et al.*, 2018, Bergler *et al.*, 1996). The *eef* remnants are predicted to form an ORF as annotated in K-12 (NC_000913.3) which may serendipitously form a useful product or this could be an erroneous prediction. Alternatively, the presence of any regulatory RNAs may also contribute to the high conservation of *eefRABCD* gene fragments. Work by Adams *et al.* has suggested activity at a transcriptional start site in this location (Adams *et al.*, 2021).

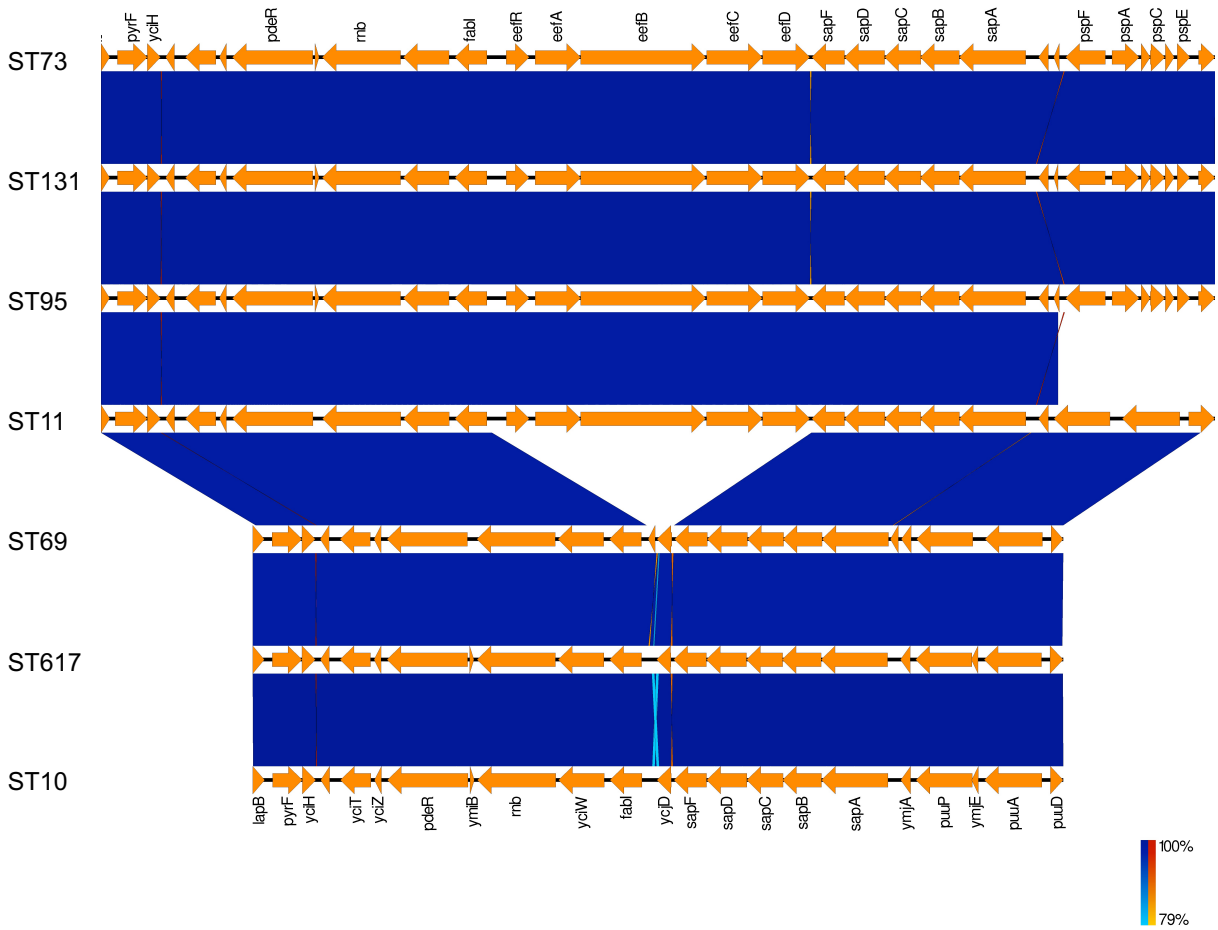


Figure 6.2 The genomic region of strains with and without *eefRABCD*

The genomes of seven annotated *E. coli* Reference Sequences were downloaded from NCBI and aligned. Maximum BLAST identity is show in navy blue, with minimum BLAST identity in light blue. Any genome inversions are shown in red in the alignments, here red indicates maximum identity and orange minimum. Four sequences belonged to STs found to possess the *eef* operon (ST73, ST131, ST95, ST11) while three did not have the operon (ST69, ST617, ST10). Where *eef* was present it was found consistently between *fabI,* which encodes an enoyl carrier critical for fatty acid biosynthesis, and *sapF* which is an putrescine transporter shown to also reduce antimicrobial peptide susceptibility. Where the *eef* operon was absent, a hypothetical

protein *ycjD* was present in place of the *eef* operon. Regardless of *eef* presence or absence, the genome region was highly conserved between STs.

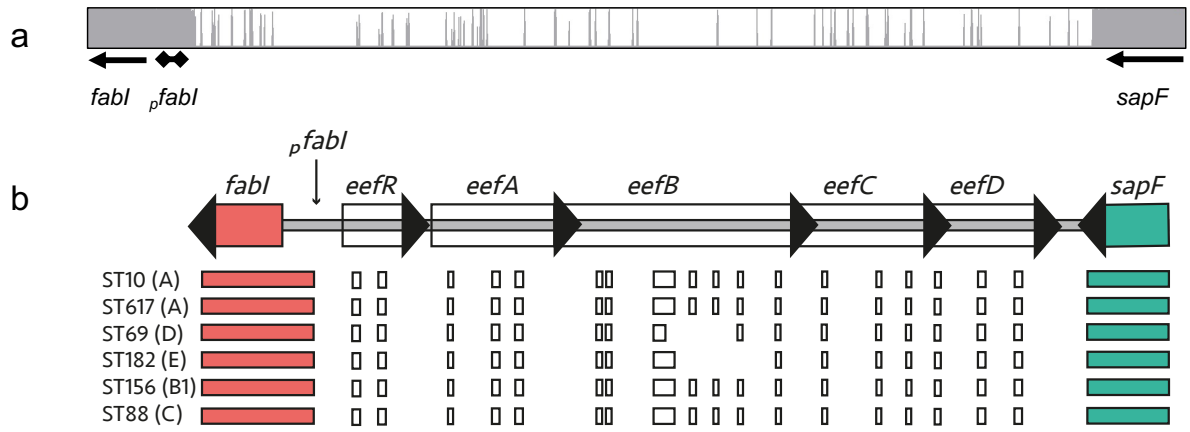


Figure 6.3 Representative homology to *eef* operon in *E. coli* STs that do not possess the operon

The *eefRABCD* operon was located between *fabI* and *sapF* in *E. coli* STs that possessed the operon, therefore the *fabI-sapF* region from *E. coli* ATCC 25922 was aligned to the *fabI-sapF* region of all non-*eefRABCD* encoding *E. coli* genomes used to construct the phylogenetic tree (Figure 6.1). This was done to determine whether fragments of *eef* remained. (a) Despite the genomic diversity of STs that lacked the *eef* operon, each ST maintained specific regions of *eefRABCD.* Grey indicates homology between the ATCC 25922 reference and the sequences that lack the *eef* operon. Both *fabI* and *sapF* were highly conserved in all sequences however regions of homology were seen between the ATCC 25922 sequence and the *fabI-sapF* intergenic region. (b) Across phylogroups A, B1 and C the same remnants remained of the *eef* operon. ST69 and ST182 had lost more of the operon but in independent patterns, suggesting *eef* had been lost from phylogroups AB1C, ST69 and ST182 in three independent events.

6.8 Genes that co-occur with the *eefRABCD* operon in *E. coli*

The *eefRABCD* operon was present in phylogroups B2, D, E, F and G but completely absent from phylogroup A. B1 and C. Due to this clear divide between phylogroups that have the *eef* operon and those that do not, it was hypothesised that *eef* may co-occur with genes important with the lifestyle associated with ExPEC. Using the same *E. coli* assemblies used to generate the phylogenetic tree, a pangenome was generated using Roary without *E. fergusonii* that was previously used as a root. Genes that were found coincident with *eef* operon across *E. coli* were determined using Coinfinder (Whelan *et al.*, 2020) (Figure 6.4 - Figure 6.8). Coinfinder first removes high and low abundance genes, genes present in < 5% of genomes as well as core genes. Once significant coincident gene pairs have been identified, the *D*-value of each gene is evaluated. *D* is a phylogenetic measure of binary traits (Fritz and Purvis, 2010, Whelan *et al.*, 2020). If a gene has a *D*-value of 1, it suggests the conservation of the gene is random, a *D* value of 0 indicates distribution of the gene is as expected under Brownian motion. However a value <0 implies a phylogenetic signal and that a gene is phylogenetically conserved (Fritz and Purvis, 2010). Roary assigns sequences to clusters based upon their translated protein sequences and assigns names either from the inputted Prokka files or incrementally, a reference proteome was therefore used to identify gene names for the co-occurring genes. This reference proteome belonged to *E. coli* O157:H7 str. Sakai (UP000000558) which is an ST11 and phylogroup E, and should therefore encode EefRABCD.

Each *eef* gene was found to co-occur with every other gene on *eefRABCD* operon supporting the hypothesis that across *E. coli* strains either possesses the whole operon or nothing at all. Furthermore, the *eef* genes were also found to co-occur with

flanking genes, *fabI* and *sapF.* Many genes found to co-occur with the *eef* operon were linked to heme catabolism (LaMattina *et al.*, 2017). The anaerobilin reductase *chuY,* ferric citrate ABC transporter permease *fecD* (two alleles across pangenome), hemin important ATO-binding protein *hmuV* and heme utilisation carrier protein *hutX* were all identified. As was the flagellar biosynthetic protein *flhB* implicated in protein export regulations (Meshcheryakov *et al.*, 2013). Moreover, a NADH-dependent flavin oxidoreductase *E3149_00275* and cytotoxic protein *ccdB* were identified (Table 6.5).

Table 6.5 Genes found to co-occur with the *eef* operon and their respective function

				eef operon				
Roary cluster	**Gene name**	***E. coli* ST11 gene name**	***E. coli* ST11 gene description**	***eefR***	***eefA***	***eefB***	***eefC***	***eefD***
LLGKEHJH_01406	*group_11399*	*chuY*	Anaerobilin reductase					
LLGKEHJH_03406	*group_12889*	*E3149_00275*	NADH-dependent flavin oxidoreductase					
LLGKEHJH_04529	*cirA_1*	*E3149_07315*	TonB dependent outer membrane receptor					
LLGKEHJH_04527	*group_14735*	*E3149_07330*	Methyltransferase					
LLGKEHJH_04526	*group_19516*	*E3149_07335*	ABC transporter ATP-binding protein					
LLGKEHJH_04524	*btuF_1*	*E3149_07345*	ABC transporter ATP-binding protein					
LLGKEHJH_01409	*hmuT*	*E3149_21545*	ABC transporter substrate-binding protein					
LLGKEHJH_01408	*group_10284*	*E3149_21550*	Coproporphyrinogen oxidase					
LLGKEHJH_00897	*salL*	*E3149_23480*	Acetolactate synthase					
LLGKEHJH_03148	*scrY*	*E3149_23950*	Glycoprotein					
LLGKEHJH_01292	*ccdA_1*	*E3149_26740*	Antitoxin					
LLGKEHJH_01291	*ccdB_1*	*E3149_26745*	Cytotoxic protein CcdB					
LLGKEHJH_03411	*yvgN_1*	*ECs_0335*	Oxidoreductase					
LLGKEHJH_00724	*mexA*	*eefA*	Efflux RND transporter periplasmic adaptor subunit					
LLGKEHJH_00723	*mexB*	*eefB (acrB_2)*	Efflux pump membrane transporter					
LLGKEHJH_00722	*mepC*	*eefC*	Outer membrane channel protein					
LLGKEHJH_00721	*bcr_2*	*eefD*	Bcr/CflA family efflux transporter					
LLGKEHJH_00725	*yfiR*	*eefR*	TetR/AcrR family transcriptional regulator					
LLGKEHJH_00726	*group_13789*	*fabI*	Enoyl-[acyl-carrier-protein] reductase [NADH] FabI					
LLGKEHJH_04525	*group_16762*	*fecD*	Ferric citrate ABC transporter permease					
LLGKEHJH_01405	*hmuU*	*fecD*	Ferric citrate ABC transporter permease					
LLGKEHJH_03256	*group_6783*	*flhB*	Flagellar biosynthetic protein FlhB					
LLGKEHJH_00729	*group_12247*	*gmr*	Cyclic di-GMP phosphodiesterase					
LLGKEHJH_01404	*hmuV*	*hmuV*	Hemin import ATP-binding protein HmuV					
LLGKEHJH_01407	*group_14662*	*hutX*	Heme utilization carrier protein					
LLGKEHJH_04528	*nadC_1*	*modD*	Putative pyrophosphorylase ModD					
LLGKEHJH_00718	*group_13786*	*sapC*	Antimicrobial peptide transport ABC transporter permease					
LLGKEHJH_00720	*sapF*	*sapF*	Antimicrobial peptide ABC transporter ATPase					
LLGKEHJH_01296	*group_9982*	*yaaU*	MFS sugar transporter					
LLGKEHJH_02278	*group_12183*	*yicJ*	Glycoside-pentoside-hexuronide family transporter					

Green indicates co-occurrence with the *eef* gene.

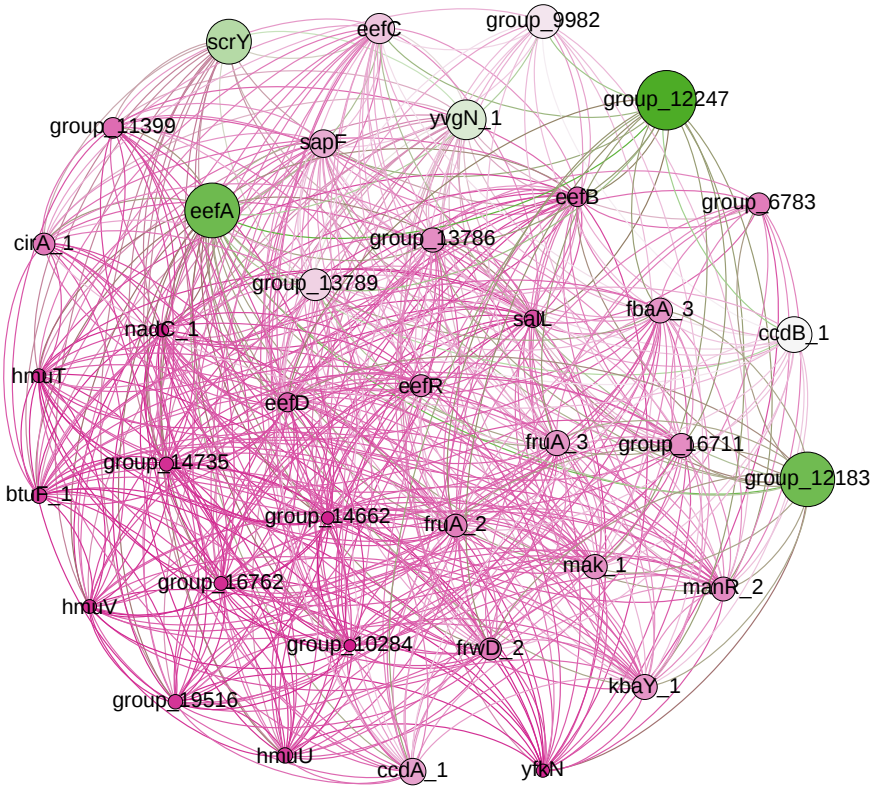


Figure 6.4 Gene network of clusters that co-occur with *eefR*

Gene clusters found to be significantly coincident with *eefR* were determined using Coinfinder and are show above as a network (Whelan *et al.*, 2020). Connections between each gene (node) and its associated genes (edges) are shown; *eefR* is shown in the centre and is connected each of the genes in the *eef* operon as well as the flanking genes, *fabI* (group_13789) and *sapF*. The size of the node circle directly correlates with *D*-value which is a phylogenetic measure of a binary trait. The smaller the node size, the smaller the *D-value* and therefore the more phylogenetically conserved the gene is. As *D*-value increases, node colour progresses from pink to green.

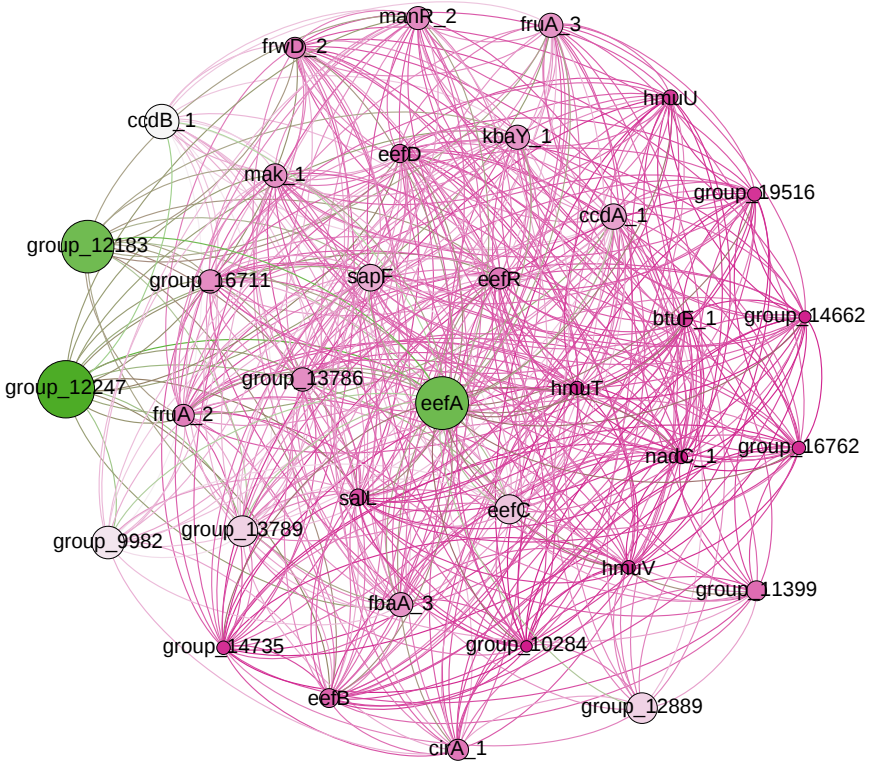


Figure 6.5 Gene network of clusters that co-occur with *eefA*

Gene clusters found to be significantly coincident with *eefA* were determined using Coinfinder and are show above as a network (Whelan *et al.*, 2020). Connections between each gene (node) and its associated genes (edges) are shown; *eefA* is shown in the centre and is connected each of the genes in the *eef* operon as well as the flanking genes, *fabI* (group_13789) and *sapF*. The size of the node circle directly correlates with *D*-value which is a phylogenetic measure of a binary trait. The smaller the node size, the smaller the *D-value* and therefore the more phylogenetically conserved the gene is. As *D*-value increases, node colour progresses from pink to green. Of the *eef* genes, *eefA* was found to have the highest *D*-value.

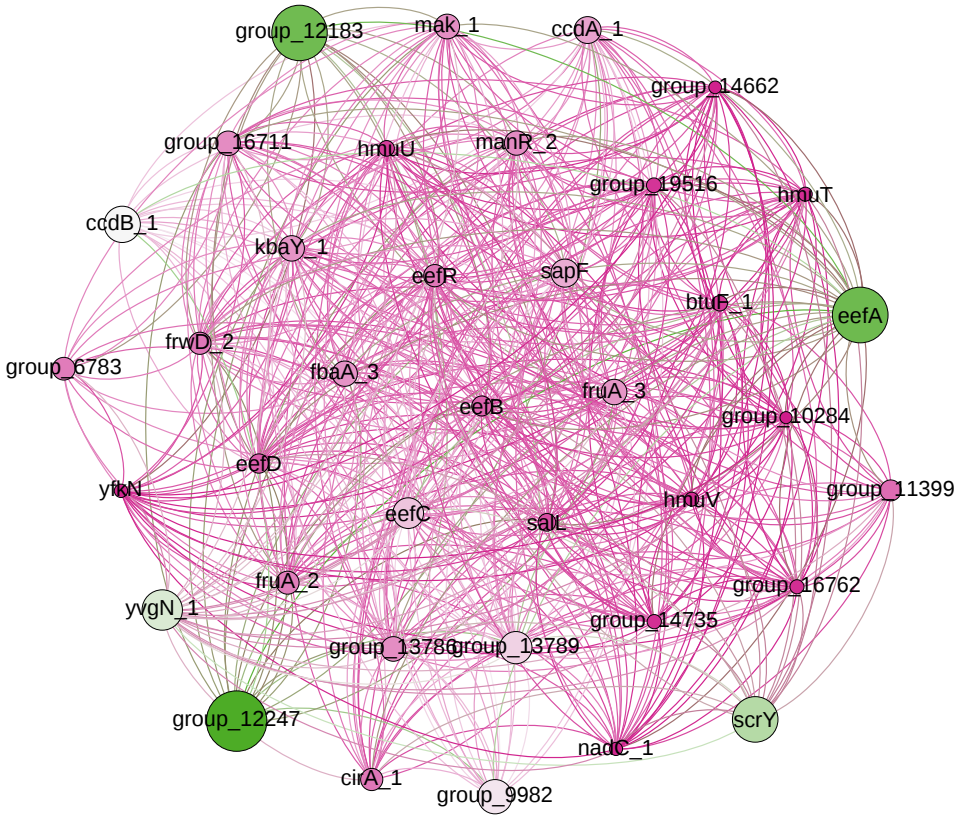


Figure 6.6 Gene network of clusters that co-occur with *eefB*

Gene clusters found to be significantly coincident with *eefB* were determined using Coinfinder and are show above as a network (Whelan *et al.*, 2020). Connections between each gene (node) and its associated genes (edges) are shown; *eefB* is shown in the centre and is connected each of the genes in the *eef* operon as well as the flanking genes, *fabI* (group_13789) and *sapF*. The size of the node circle directly correlates with *D*-value which is a phylogenetic measure of a binary trait. The smaller the node size, the smaller the *D-value* and therefore the more phylogenetically conserved the gene is. As *D*-value increases, node colour progresses from pink to green.

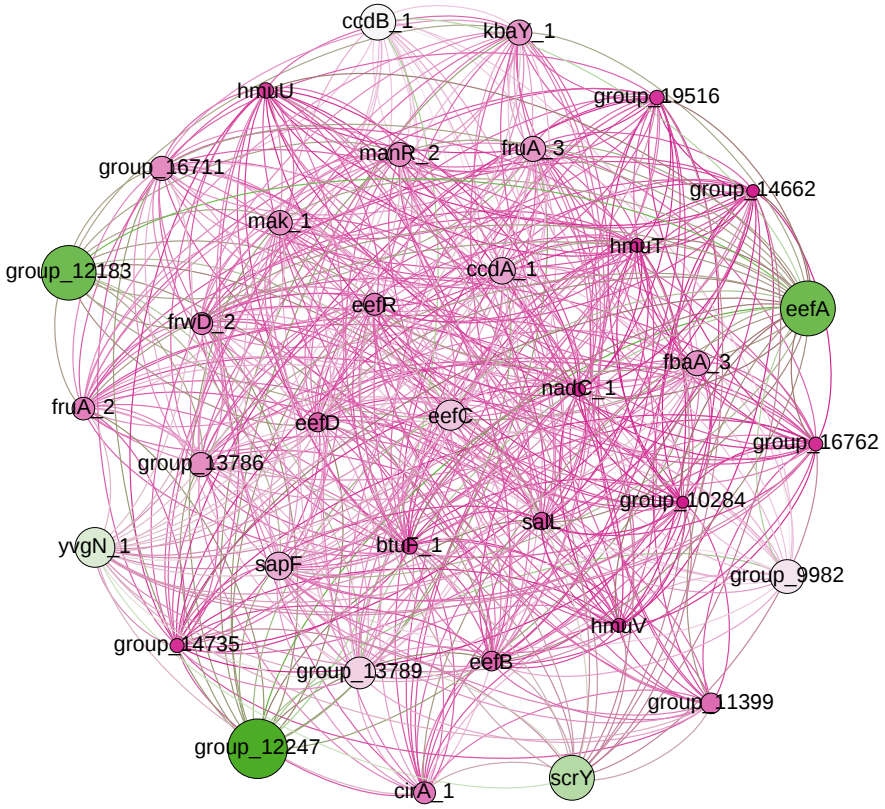


Figure 6.7 Gene network of clusters that co-occur with *eefC*

Gene clusters found to be significantly coincident with *eefC* were determined using Coinfinder and are show above as a network (Whelan *et al.*, 2020). Connections between each gene (node) and its associated genes (edges) are shown; *eefC* is shown in the centre and is connected each of the genes in the *eef* operon as well as the flanking genes, *fabI* (group_13789) and *sapF*. The size of the node circle directly correlates with *D*-value which is a phylogenetic measure of a binary trait. The smaller the node size, the smaller the *D-value* and therefore the more phylogenetically conserved the gene is. As *D*-value increases, node colour progresses from pink to green.

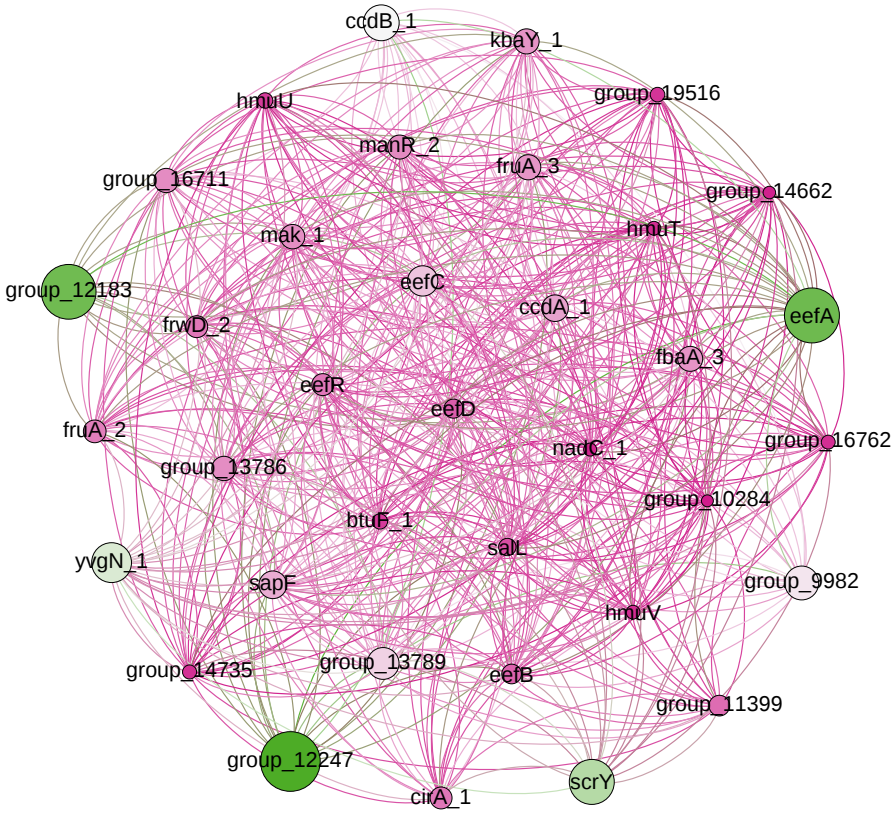


Figure 6.8 Gene network of clusters that co-occur with *eefD*

Gene clusters found to be significantly coincident with *eefD* were determined using Coinfinder and are show above as a network (Whelan *et al.*, 2020). Connections between each gene (node) and its associated genes (edges) are shown; *eefD* is shown in the centre and is connected each of the genes in the *eef* operon as well as the flanking genes, *fabI* (group_13789) and *sapF*. The size of the node circle directly correlates with *D*-value which is a phylogenetic measure of a binary trait. The smaller the node size, the smaller the *D-value* and therefore the more phylogenetically conserved the gene is. As *D*-value increases, node colour progresses from pink to green. The *eefD* was found to have the smallest *D*-value.

6.9 Modelling suggests EefABC is most similar to RND systems from other organisms

To try to understand the structure of EefABC and determine how similar the system is to the other known RND pumps within *E. coli,* protein homology modelling was done by Dr. Vassiliy Bavro (Essex University). This confirmed that EefA, EefB and EefC do indeed form a tripartite RND efflux complex (Figure 6.9). However, the pump is predicted to have unusual features compared to other *E. coli* RND systems. Compared with other *E. coli* efflux proteins EefC, the OMP component of EefABC, was found to be most similar to CusC of the copper and silver efflux pump. However, comparing the protein to a wider range of OMPs it became apparent that the structure of EefC is most similar to OprM/OprJ from *Pseudomonas* spp. The RND pump EefB was most structurally similar to AcrB and MexB of *E. coli* and *P. aeruginosa,* respectively, while the PAP protein EefA was most similar to MexA (*P.* aeruginosa) and AcrA (*E. coli*). Analysis of the overall EefABC system showed close similarities between MexAB-OprM and MexCD-OprJ. One explanation for this would be that the pump had been acquired via horizontal gene transfer. However, in *E. coli* ATCC 25922, which belongs to ST73 (phylogroup B2), the GC percentage across the *eef* operon is 51%, which is exactly the same as the average GC percentage across the genome in this strain suggesting that the *eefRABCD* operon has long been a part of the *E. coli* genome. Thus, due to the high level of conservation across phylogroups and minimal GC content differences between the *eef* operon and wider genome of *E. coli* it was hypothesised that EefABC was a pump either originating in *E. coli* or acquired prior to the divergence of *E. coli* phylogroups and has evolved over time.

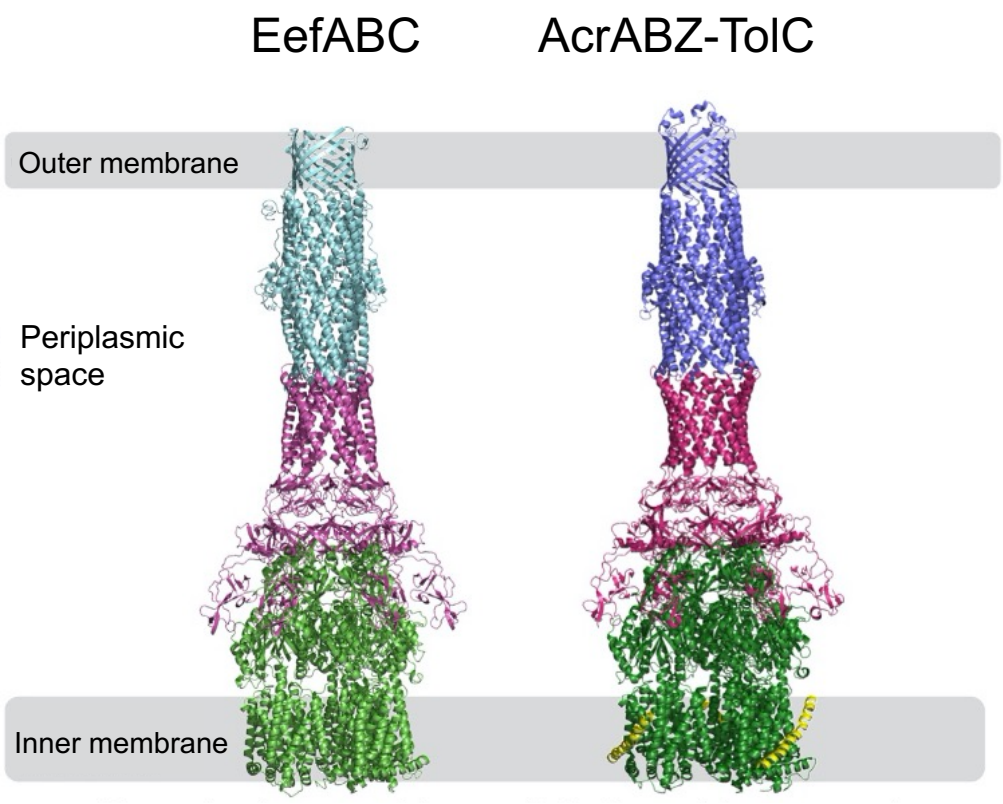


Figure 6.9 Predicted structure of EefABC compared to AcrAB-TolC

The predicted structure of EefABC shown on the left in comparison with the experimentally derived cryo-EM structure of AcrABZ-TolC (PBD: 5066) (Wang *et al.*, 2017). Modelled by Dr. Vassiliy Bavro.

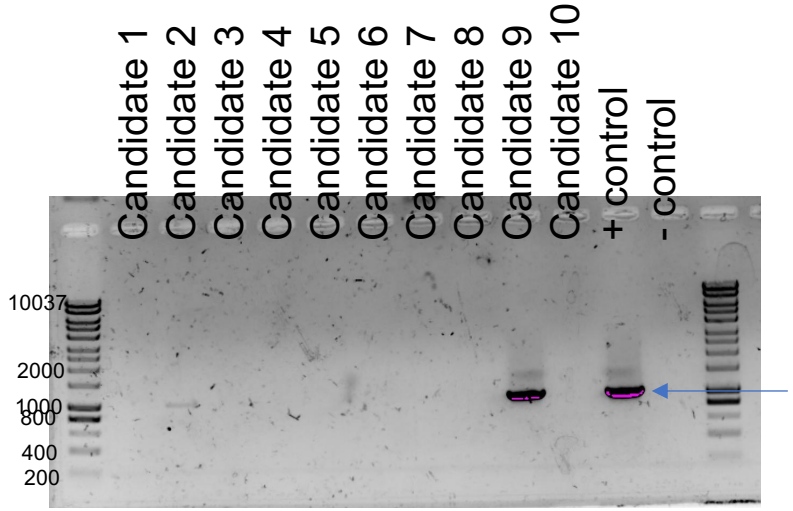
6.10 Functional characterisation of EefABC

6.10.1 Selection of a laboratory strain for *eef* work

The only *E. coli* strain in which the *eefRABCD* operon has been previously described is *E. coli* SMS-3-5 (Colclough *et al.*, 2019, Fricke *et al.*, 2008). *E. coli* SMS-3-5 was isolated from a heavily polluted coastal region in North America and has a high tolerance to five classes of antimicrobials, therefore an alternative strain was desired for any future laboratory characterisation experiments to aid genetic manipulation as the presence of multiple resistances would prevent the use of common genetic markers. As mentioned, one of the strains identified in the *eefB* BLASTN across *E. coli* was ATCC 25922. This strain is an MIC reference strain with known antimicrobial susceptibilities and available from Public Health England. This made the strain more appealing for any future laboratory work and therefore the nucleotide and protein similarity with SMS-3-5 were determined. SMS-3-5 and ATCC 25922 have 99% sequence identity to one another with 52 nucleotides differing out of 7493. This translated into only 5 amino acid substitutions at the protein level (Appendix 8.3). Two in EefA, two in EefC and one in EefD. Notably the regulator EefR and the EefB RND pump were identical at the amino acid level.

6.10.2 EefABC is not a drug transport pump

To identify any substrates of the RND pump EefABC, *eefABC* from *E. coli* ATCC 25922 was cloned into the pET21a and pET24a vectors. These vectors only differ by the selective marker, pET21a carries an ampicillin resistance marker and pET24a a kanamycin resistance marker. This allowed the susceptibility profiles of all classes of antibiotics to be determined. Briefly the *eefABC* operon was amplified from ATCC 25922. Both plasmids and the PCR product were digested using *Nde*I and *Xho*I. Following ligation the final plasmid sizes were 10,973 and 10,907 for pET21a (Figure 6.10) and pET24a (Figure 6.11) respectively. Inserts were confirmed by sequencing.



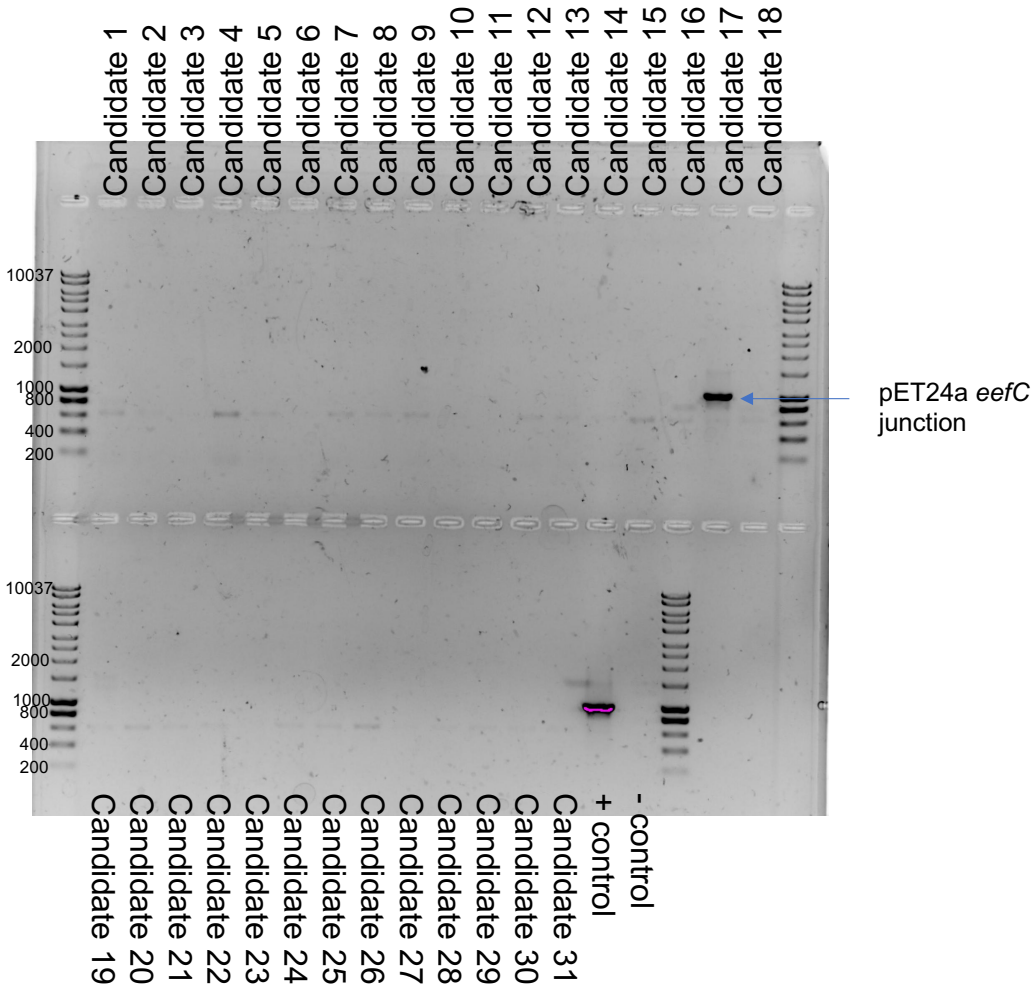
pET21a *eefC* junction

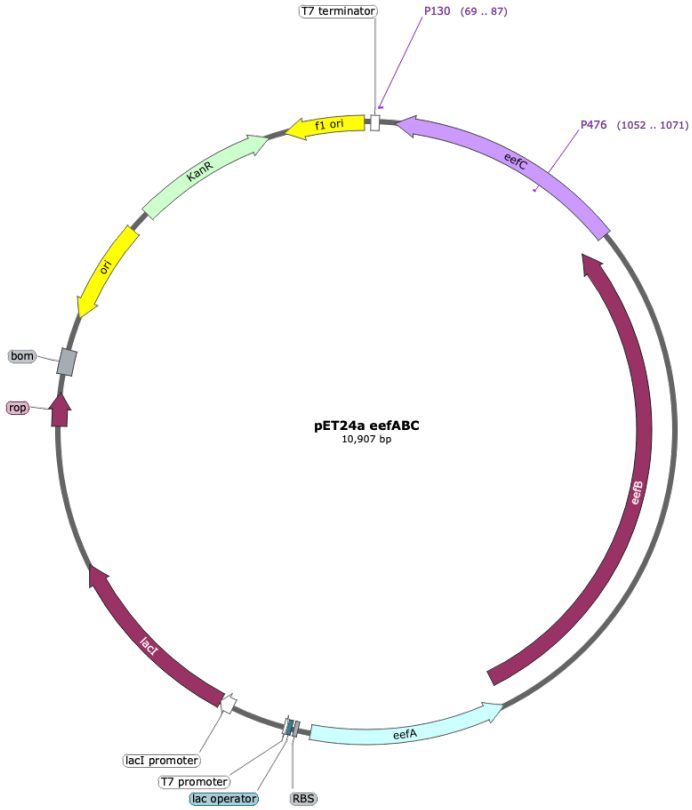


Expected product size 1003 bp

Figure 6.10 Confirmation of pET21a *eefABC* cloning

As the *eefABC* insertion was a similar size to the pET21a backbone, pET21a *eefABC* candidates were confirmed by PCR and NGS sequencing using a primer that annealed internally within *eefC* (P476) and another that bound to the plasmid backbone (P130). A previously cloned and sequenced pET21a *eefABC* construct that encoded a T7 signal sequence was used as a positive control, while the negative control was a PCR reaction that contained no DNA template. Successful candidates with an amplimer size of ~1000 bp were purified and sequenced.





Expected product size 1003 bp

Figure 6.11 Confirmation of pET24a *eefABC* cloning

As the *eefABC* insertion was a similar size to the pET24a backbone, pET21a *eefABC* candidates were confirmed by PCR and NGS sequencing using a primer that annealed internally within *eefC* (P476) and another that bound to the plasmid backbone (P130). A previously cloned and sequenced pET24a *eefABC* construct that encoded a T7 signal sequence was used as a positive control, while the negative control was a PCR reaction that contained no DNA template. Successful candidates with an amplimer size of ~1000 bp were purified and sequenced.

Strains with *eefB* and *eefABC* gene inactivations were constructed using homologous recombination. Initial attempts to inactivate the *eefABC* operon using one-step homologous recombination were unsuccessful. It was hypothesised this was due to differences in DNA fragment lengths between *eefABC::aph* and *eefABC* on the chromosome. As a *result eefABC* inactivation was done in a two-step process (Figure 6.12). Initially the RND gene *eefB* was interrupted through the introduction of the *aph* gene which was subsequently removed using pCP20. Following confirmation of this first gene deletion by sequencing the remaining components of the *eefABC* operon, *eefA* and *eefC,* were inactivated in a second step using homologous recombination (Figure 6.13).

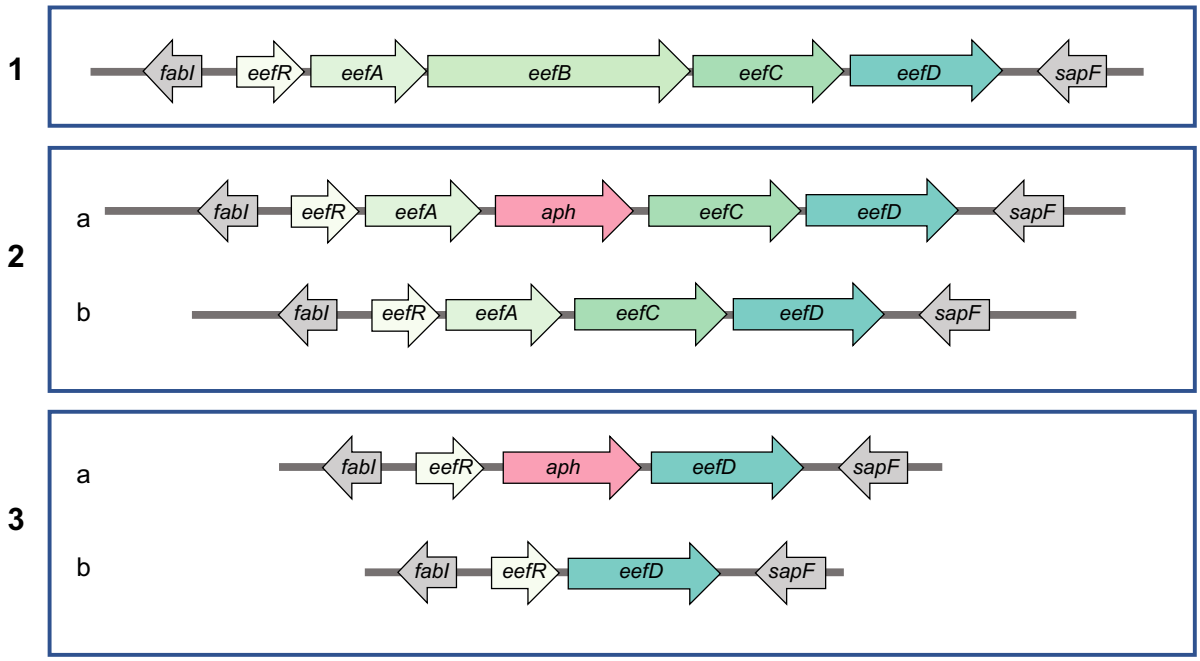


Figure 6.12 Approach used to inactivate *eefABC*

The *eefABC* operon (1) was first interrupted through the introduction of an *aph* cassette into the *eefB* which encodes an RND efflux pump (2a). This *aph* cassette was then removed using pCP20 treatment (2b). Next an *aph* cassette was used to interrupt *eefA* and *eefC* in one step (3a), this *aph* cassette was once again removed using pCP20 treatment (3b). This approach resulted in the construction of both Δ*eefB* and Δ*eefABC* strains.

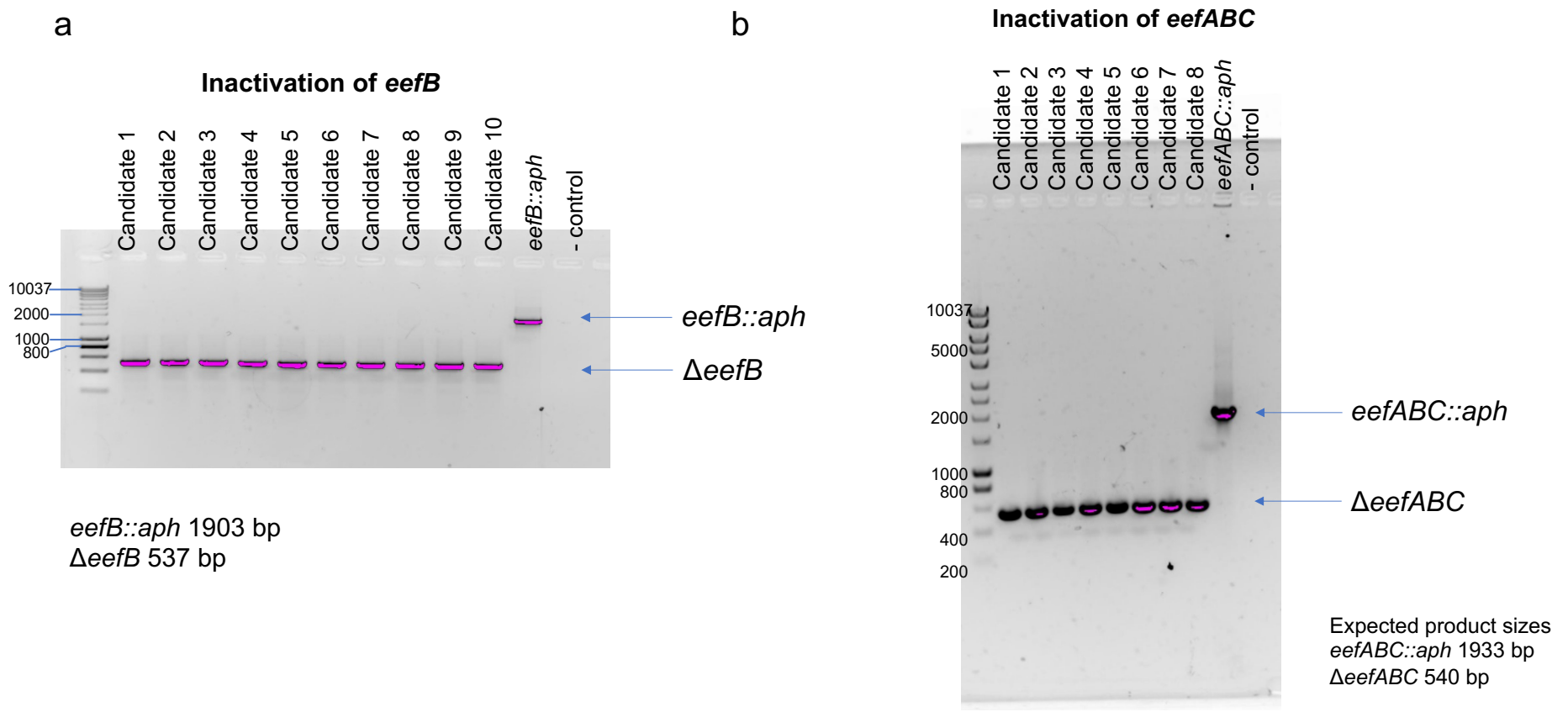


Figure 6.13 Confirmation of *eefB* and *eefABC* gene inactivation

Introduction and removal of the *aph* cassette were confirmed via PCR and NGS. a) Confirmation of *eefB* interruption via the *aph* cassette was confirmed via PCR amplification using P413 and P414 which annealed outside the gene. Introduction of the *aph* cassette

reduced the amplimer size from 3535 bp (*E. coli* ATCC 25922) to 1903 bp (*E. coli* ATCC 25922 *eefB::aph*). Removal of the *aph* cassette further reduced the amplimer to 537 bp. b) Interruption of the remaining *eefABC* operon was confirmed by the amplification of the *eefABC* region using P415 and P416 that annealed flanking DNA. Introduction of the *aph* cassette into the *eef* operon for a second time reduced the amplimer size from 3041 bp (*E. coli* ATCC 25922 Δ*eefB*) to 1933 bp (*E. coli* ATCC 25922 *eefABC::aph*). Further removal of the *aph* cassette reduced the amplimer size to 540 bp.

RND efflux pumps have been demonstrated to confer MDR in both laboratory and clinical settings, thus it was hypothesised that EefABC would also export antimicrobials. To test this the antimicrobial susceptibility of strains with *eefABC* inactivated was measured along with strains expressing *eefABC* that do not normally encode it.

Inactivation of the RND pump, *eefB* did not result in any increases to drug susceptibility (Table 6.6), nor did further inactivation of the remaining *eefABC* genes, *eefA* and *eefC*. Furthermore introduction of *eefABC* into K-12 did not reduce susceptibility to any of the antimicrobials tested (Table 6.6). AcrB has a wide substrate range and is expressed under laboratory conditions and can therefore mask subtle susceptibility changes conferred by additional RND efflux pumps. As a result, *eefABC* was expressed in K-12 Δ*acrB*.

Deletion of *acrB* significantly increased susceptibility to almost all antimicrobials tested. However expression of *eefABC* did not result in any significant reduction of susceptibility to antimicrobials (Table 6.6). A single fold increase from 8 μg/mL to 16 μg/mL was observed for the MIC of fusidic acid when *eefABC* was expressed on both the pET21a and pET24a plasmid. Single fold increases were also observed for benzalkonium chloride and novobiocin when *eefABC* was expressed in the K-12 Δ*acrB* background.

In addition to antimicrobials, the MICs of four dyes (crystal violet, ethidium bromide, methylene blue, rhodamine 6G) were also measured. Inactivation of *eefB* and *eefABC* had no effect on dye susceptibility (Table 6.7). Furthermore expression of *eefABC* in the K-12 background did not further reduce dye susceptibility. Inactivation of the RND transporter of *acrB* significantly increased susceptibility all four dyes,

however expression of *eefABC* in the K-12 Δ*acrB* background reduced susceptibility to ethidium bromide and rhodamine 6G from 8 μg/mL to 32 μg/mL.

The *eefABC* operon was expressed using both the pET21a and pET24a vectors, the former encodes an ampicillin resistance marker and the latter kanamycin. This allowed the susceptibility of all antimicrobial classes for strains expressing the *eefABC* operon to be determined. Using both vectors the same patterns were seen suggesting *eefABC* does indeed reduce susceptibility to ethidium bromide and rhodamine 6G.

Table 6.6 Susceptibility of *E. coli* to antimicrobials following loss and gain of *eefABC* and *eefD*

	MIC (μg/mL)																
Strain	**AZT**	**BAC**	**CAR**	**CEF**	**CHL**	**CLI**	**ERY**	**FA**	**GEN**	**MER**	**MOX**	**NA**	**NOV**	**RIF**	**SPE**	**TET**	**TIC**
ATCC 25922	0.12	32	16	0.06	4	128	64	512	0.5	0.016	0.03	2	32	>32	8	1	8
ATCC 25922 Δ*eefB*	0.12	32	16	0.06	4	128	64	512	0.5	0.016	0.03	2	32	>32	8	1	8
ATCC 25922 Δ*eefABC*	0.25	32	16	0.06	2	128	64	512	0.5	0.016	0.03	2	32	>32	8	1	8
MG1655 + pET21a	0.12	64	>1024	0.06	8	128	32	512	0.12	0.03	0.06	8	128	>32	4	1	>1024
MG1655 + pET21a *eefABC*	0.12	64	>1024	0.06	8	128	32	1024	0.25	0.03	0.06	8	128	>32	4	1	>1024
MG1655 + pET24a	0.06	64	8	0.06	8	128	32	512	0.25	0.03	0.03	8	128	>32	8	1	4
MG1655 + pET24a *eefABC*	0.06	64	8	0.06	8	128	32	512	0.25	0.016	0.06	8	128	>32	4	1	4
MG1655 Δ*acrB* + pET21a	0.12	4	>1024	<0.016	1	2	2	8	0.12	0.03	<0.008	2	2	32	4	0.25	>1024
MG1655 Δ*acrB* + pET21a *eefABC*	0.06	8	>1024	0.03	1	2	2	16	0.12	0.03	<0.008	2	4	32	4	0.25	>1024
MG1655 Δ*acrB* + pET24a	0.12	4	8	<0.016	1	2	2	8	0.25	0.03	<0.008	2	2	32	4	0.25	4
MG1655 Δ*acrB* + pET24a *eefABC*	0.06	8	8	<0.016	2	2	4	16	0.25	0.03	<0.008	2	4	32	4	0.25	4
ATCC 25922	0.12	32	16	0.06	4	128	64	512	0.5	0.016	0.03	2	32	>32	8	1	8
ATCC 25922 Δ*eefD*	0.12	32	16	0.06	8	256	64	1024	0.5	0.016	0.06	8	64	>32	8	1	8
MG1655 + pACYC177	0.06	64	>1024	0.06	8	128	32	1024	0.5	0.03	0.06	8	128	>32	4	1	>1024
MG1655 + pACYC177 *eefD*	0.06	64	8	0.06	8	128	32	1024	0.5	0.03	0.06	8	128	>32	4	1	4
MG1655 Δ*acrB* + pACYC177	0.12	4	>1024	<0.016	1	2	2	8	0.5	0.03	<0.008	2	2	32	4	0.25	>1024
MG1655 Δ*acrB* + pACYC177 *eefD*	0.12	4	8	<0.016	1	2	2	8	0.5	0.03	<0.008	2	2	32	4	0.25	4

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

Table 6.7 Susceptibility of *E. coli* to dyes following loss and gain of *eefABC* and *eefD*

	MIC (μg/mL)			
Strain	**CV**	**EB**	**MB**	**R6G**
ATCC 25922	32	512	>1024	1024
ATCC 25922 Δ*eefB*	32	512	>1024	512
ATCC 25922 Δ*eefABC*	32	512	>1024	512
MG1655 + pET21a	16	1024	>1024	512
MG1655 + pET21a *eefABC*	32	1024	>1024	512
MG1655 + pET24a	16	1024	>1024	512
MG1655 + pET24a *eefABC*	16	1024	>1024	512
MG1655 Δ*acrB* + pET21a	4	8	8	8
MG1655 Δ*acrB* + pET21a *eefABC*	4	**32**	16	**32**
MG1655 Δ*acrB* + pET24a	4	8	8	8
MG1655 Δ*acrB* + pET24a *eefABC*	4	**32**	16	**32**
ATCC 25922	32	512	>1024	1024
ATCC 25922 Δ*eefD*	32	512	>1024	1024
MG1655 + pACYC177	32	1024	>1024	512
MG1655 + pACYC177 *eefD*	32	1024	>1024	512
MG1655 Δ*acrB* + pACYC177	4	8	8	8
MG1655 Δ*acrB* + pACYC177 *eefD*	4	8	8	8

CV – crystal violet, EB – ethidium bromide, MB – methylene blue, R6G – rhodamine 6G.

6.11 Characterisation of the associated MFS pump, *eefD*

Due to the proximity and high conservation of *eefD,* this MFS pump was also deleted although this was independent to the inactivation of *eefABC.* As the promoter of *eefD* was unknown, the gene was cloned into the pACYC177 plasmid, interrupting the ampicillin resistance gene to make use of this promoter. Both the deletion of *eefD* from ATCC 25922 and the introduction of *eefD* to K-12 Δ*acrB* were found to have no effect on the MICs of any antimicrobials (Table 6.6) or dyes tested (Table 6.7).

6.12 Presence or absence of *eefRABCD* in other Gram negative bacterial species

The EefABC efflux pump was first described in *K. aerogenes* but has also been identified in *K. pneumoniae,* though the operons present differ between the two *Klebsiella* species (Coudeyras *et al.*, 2008, Masi *et al.*, 2005). In *K. aerogenes* the operon consists of only *eefABC* yet in *K. pneumoniae* there is an MFS pump downstream of *eefC* as seen in *E. coli* (Figure 6.14). Neither *K. pneumoniae* nor *K. aerogenes* were found to possess *eefR.*

The presence of *eefRABCD* across the *Escherichia* genus was determined using BLASTN. The *eef* operon from ATCC 25922 was used to search *Escherichia* (Taxid 561) nucleotide collection, with *E. coli* (Taxid 562) sequences excluded. Only *Escherichia marmotae* and *Escherichia albertii* were found to possess the *eefRABCD* operon and as seen in *E. coli* the five gene operon *eefRABCD* was located between *fabI* and *sapF* (Figure 6.15). *Escherichia alba, E. fergusonii, Escherichia ruysiae,* and *Escherichia whittamii* were not found to have operon. Due to the occurrence of *eef* genes across both *Escherichia* and *Klebsiella*, its presence in closely related Gammaproteobacteria was determined.

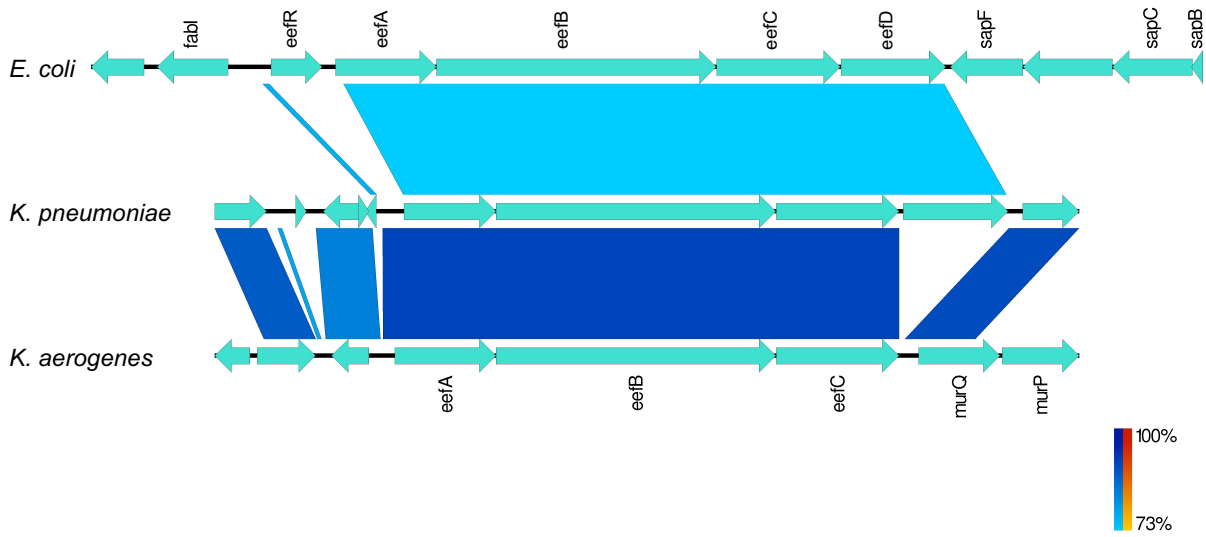


Figure 6.14 Alignment of the *E. coli* and *Klebsiella* spp. *eef* operons

The presence of *eefABC* has been previously described in both *K. pnuemoniae* (NC_016845.1) and *K. aerogenes* (NZ_CP041925.1). A representative genome of each *Klebsiella* species was therefore selected and the *eef* operons of both speices were aligned to that of *E. coli* ATCC 25922 (CP_009072.1). Alignment confimed that *eefABC* are found in both *K. pnuemoniae* and *K. aerogenes,* however *eefD* is only present in *K. pneumoniae.* Moreover both *Klebsiella* spp. lack the *eefR*.

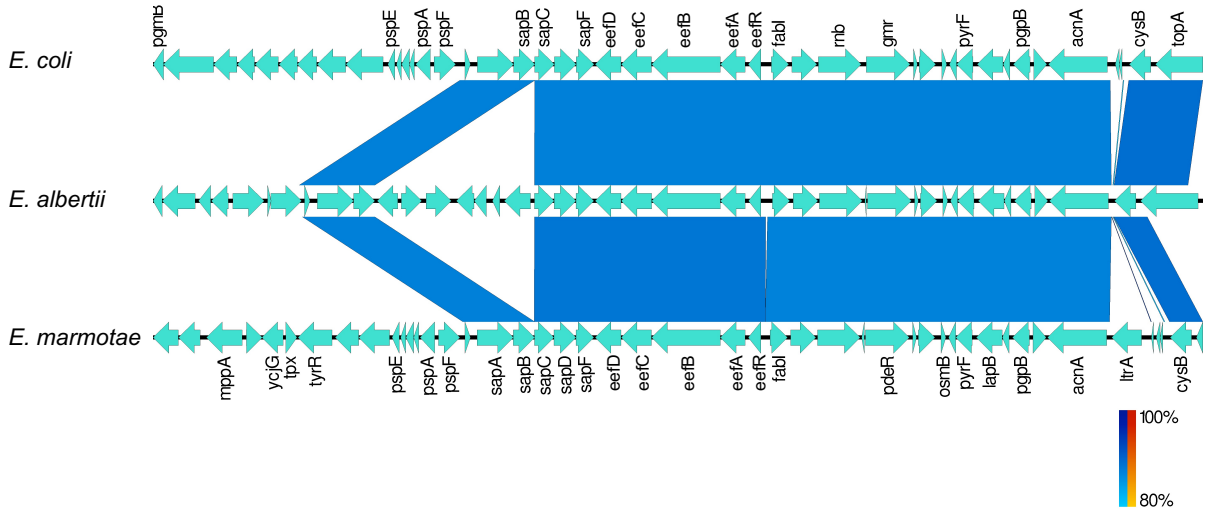


Figure 6.15 Alignment of the *eef* operon of multiple *Escherichia* spp.

As *eefRABCD* was found to be highly prevalent across phylogroups B2, D, E, F and G of *E. coli,* the presence of the *eef* operon across the *Escherichia* genus was determined (excluding *E. coli*). Nucelotide sequences belonging to *E. albertii* and *E. marmotae* were found to also possess the operon. *E. coli* ATCC 25922 (CP_009072.1) was subsequently aligned to *E. albertii* (CP070290.2) and *E. marmotae* (CP056165.1) confirming the presence of *eefRABCD.* Furthermore, the *eef* operon in these two *Escherichia* spp. is found between *fabI* and *sapF* as in *E. coli.*

Due to the genetic similarity of *Shigella* and *E. coli* (van den Beld and Reubsaet, 2012, Brenner *et al.*, 1972), the presence of *eef* across *Shigella spp.* was determined using the *E. coli* ABRicate custom database. For each of the *Shigella* spp., 500 assemblies were downloaded from Enterobase, following the removal of duplicate assemblies and ABRicate results split over multiple contigs, over 350 assemblies were used in the analysis for each species (Table 6.8). Wide spread presence of the operon was only seen in *S. dysenteriae* (Table 6.9). In total 363 *S. dysenteriae* were found to have the at least partial coverage of four *eef* genes. Gene fragments were detected also in a small number of *S. boydii* and *S. flexneri* assemblies. However, none of the *S. sonnei* assemblies were found to have any of the *eef* genes.

Table 6.8 *Shigella* assemblies used in this analysis

Species	Assemblies downloaded	Duplicates removed	Removed due to split genes	Assemblies in analysis
Shigella boydii	495	31	0	469
Shigella dysenteriae	497	11	4	486
Shigella flexneri	499	148	0	352
Shigella sonnei	500	140	0	360

Table 6.9 Presence and subsequent gene coverage of *eefRABCD* across the *Shigella* species with reference to the *E. coli* ATCC 25922 operon

		Percentage coverage				
	Genomes (*n*)	*eefR*	*eefA*	*eefB*	*eefC*	*eefD*
Shigella boydii	469	2.9851	2.9851	1.9944	1.919	2.4806
Shigella dysenteriae	486	0.3956	20.084	75.516	75.514	74.74
Shigella flexneri	352	0.0028	0.0028	0.0028	0.0028	0.0028
Shigella sonnei	360	0	0	0	0	0

Over 75% of *S. dysenteriae* assemblies were found to have *eef* genes present in their genome, yet only one assembly had all five genes (assembly ESC_CA5622AA_AS). All other assemblies positive for the *eef* operon completely lacked the regulator, *eefR* and lacked the majority of the second operon gene, *eefA*. Though gene coverage was not always 100%, the sequence identity between the reference from *E. coli* ATCC 25922 and the *Shigella* spp. was found to be greater than 98% (Table 6.10). As *eefA* follows *eefR* in the operon and both were found to have reduced coverage, this suggested that the first section of the operon may have been interrupted in *S. dysenteriae*.

Table 6.10 Conservation as percentage nucleotide identity of *eefRABCD* with reference to *E. coli* ATCC 25922 across *Shigella* species

		Percentage identity				
	Genomes containing *eef* (*n*)	*eefR*	*eefA*	*eefB*	*eefC*	*eefD*
Shigella boydii	15	99.222	97.95	96.633	97.318	97.888
Shigella dysenteriae	370	99.09	97.622	98.288	97.382	97.251
Shigella flexneri	1	99.12	97.95	98.17	97.53	98.37
Shigella sonnei	0	0	0	0	0	0

For this reason, 15 *S. dysenteriae* assemblies positive for the operon were selected at random to determine operon conservation in relation to a reference sequence from the NCBI RefSeq database, *S. dysenteriae* ATCC 12039 (CP026674). Genomes were annotated using PROKKA and aligned. Annotation predicted the presence of IS both at the start and the end of the operon. Noticeably, IS were predicted in the location of *eefR* and the start of *eefA.* Thus the 15 assemblies and reference were put through ISEscan, an IS element predictor for prokaryotic genomes (Xie and Tang, 2017). As in *E. coli*, the *eef* genes in *Shigella* spp. were found to be consistently placed between *fabI* and *sapF,* which encode an enoyl carrier and putrescine transporter, respectively however IS1 family elements were present at the start of the operon, while an IS110 element was located between *eefD* and *sapF* (Figure 6.16). The presence of IS elements in place of *eefR* and *eefA* may indicate *eefRABCD* is being lost.

The *eefRABCD* operon was found to be completely absent from *Salmonella* (Taxid 590) and *Photorhabdus* (Taxid 29487) species*,* though remnants were present in both *Yersinia* (Taxid 629) and *Serratia* (Taxid 613). Closer inspection found that both *Yersinia pestis* and *Serratia marcescens,* possessed an RND pump with 99% amino acid coverage to the peptide sequence of EefB in *E. coli* ATCC 25922. The sequence identities of these *eefB* homologs was found to be 71.95% and 68.70% for *Y. pestis* and *S. marcescens,* respectively. Despite the high number of RND efflux pumps in *A. baumannii* and *P. aeruginosa*, these two species do not possess any of the *eef* genes.

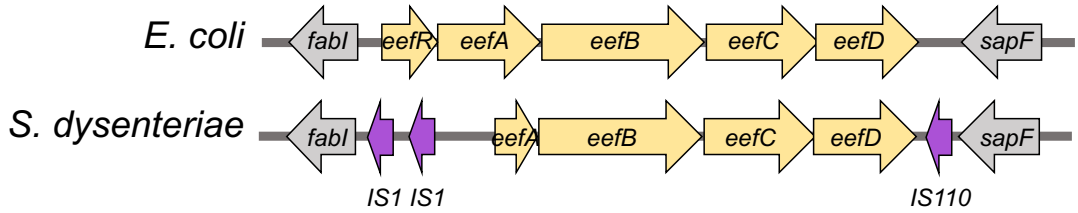


Figure 6.16 Operon comparison between *E. coli* and *S. dysenteriae*

Of the four *Shigella* spp, the *eef* operon was most prevalent across *S. dysenteriae* however the coverage seen for *eefR* and *eefA* was < 0.4% and 20.1%, respectively. Further analysis identified fragments of homology to IS elements where *eefR* and *eefA* are usually located in the operon.

6.13 Discussion

The prevalence of RND efflux pumps across clinically relevant strains of bacteria has been shown to vary. In *A. baumannii* over-expression of AdeABC and AdeIJK is associated with MDR yet the detection of *ade* genes in clinical isolates, in particular *adeB*, has been found to vary between studies (Leus *et al.*, 2020, Nemec *et al.*, 2007, Wieczorek *et al.*, 2013, Hou *et al.*, 2012, Nowak *et al.*, 2015). Despite this known variation between *A. baumannii* isolates, *E. coli* is still assumed to have six RND systems. However, in Chapter 4 it was shown that *E. coli* ST11 (O157) lacks a functional AcrF despite the virulence of this lineage and the connection between bacterial virulence and AcrF so therefore, this assumption is unjustified. In this chapter it was shown that many *E. coli* actually have seven RND pumps; AcrB, AcrD, AcrF, MdtBC, MdtF, CusA, and EefB.

RND pumps are most frequently encoded chromosomally, yet several examples of RND efflux pumps mobilised on plasmids have been published. The OqxAB pump was first identified in *E. coli* isolated from a swine farm but has since been found distributed across human clinical isolates of Enterobacteriaceae via the pOLA52 plasmid and is believed to have been mobilised from the *Klebsiella* chromosome (Hansen *et al.*, 2004) (Hong *et al.*, 2009, Li *et al.*, 2019b). Furthermore, carriage of an RND pump originating from *Pseudomonas* has been identified in *K. pneumoniae* demonstrating that the transfer of RND efflux pumps is not limited within an order but once mobilised, RND pumps can be disseminated across taxonomical orders (Lv *et al.*, 2020). Novel and previously uncharacterised systems have also been recently reported. A putative RND system, H239_3064, was discovered on the chromosome of *K. pneumoniae* and has been linked to colistin resistance (Cheng *et al.*, 2018),

demonstrating that unknown and uncharacterised RND pumps are present in Gram-negative pathogens of AMR-related significance.

EefABC was initially identified in the opportunistic human pathogen *K. aerogenes* (formerly *Enterobacter*) where it was shown to be repressed by H-NS and expressed cryptically (Masi *et al.*, 2006, Masi *et al.*, 2005). However, H-NS silencing of the *eef* operon in *E. coli* is highly unlikely due to proximity of the *eefR* promoter to that of *fabI* (Figure 6.3) as this gene has been demonstrated to be essential for *E. coli* K-12 growth via TraDIS (Goodall *et al.*, 2018). Differences between operons in *K. pneumoniae* and *K. aerogenes* have also been reported as *K. pneumoniae* possess more of the *eef* operon than *K. aerogenes*. As in *E. coli,* an MFS pump was found downstream of *eefABC* in *K. pneumoniae* but this was absent in *K. aerogenes*, yet both *Klebsiella* spp. lack the upstream TetR regulator.

The OMP of the EefABC system, EefC, has been suggested to have weak cationic selectivity when compared to TolC in *K. aerogenes,* and it has therefore been suggested that the substrates of EefC would differ to those of TolC, yet in this work gain of *eefABC* only resulted in reduced susceptibility to ethidium bromide and rhodamine 6G (Masi *et al.*, 2007). However as discussed in Chapter 5 a lack of measurable phenotype does not mean that there is no effect. The lack of detection could be because the relevant substrates were not yet tested. Furthermore, further inactivation of *acrF* may increase assay resolution as AcrF has a similar substrate range to AcrB and may therefore be masking subtle phenotypic changes. For example *acrF* inactivation may result in expression of *eefABC* having a more significant impact on benzalkonium chloride, fusidic acid and novobiocin susceptibility.

Presence of the *eefRABCD* operon was not found sporadically across the *E. coli* phylogroups and lineages but was instead highly conserved in phylogroups B2, D, E, F and G, whilst completely absent from A, B1 and C. *E. coli* isolated as the causative agent of ExPEC infections are most commonly belonging to phylogroups B2 and D, while phylogroups F and G are both associated with high virulence and AMR carriage (Clermont *et al.*, 2019, Vangchhia *et al.*, 2016). Furthermore, recent modelling of the evolution of *E. coli* suggested A, B1 and E had gained carbohydrate metabolism genes whilst phylogroups B2, F and G lost genes associated with environmental-adaptive functions and become associated with virulent lifestyles (Gonzalez-Alba *et al.*, 2019). It is therefore possible that the *eef* operon is highly important in colonisation and survival within the given niches of these lineages as RND efflux pumps are not only involved in MDR but in processes including biofilm production, quorum sensing, siderophore efflux, protection against pH, bile salts, nitrosative stress and plant derived toxins (Henderson *et al.*, 2021). Although not shown in *E. coli,* increased expression at the *eef* promoter was seen following acid shock and the deletion of *eefA* reduced the gastrointestinal colonisation of *K. pneumoniae in vivo,* supporting the hypothesis that *eef* may be most important during host colonisation and extreme conditions (Coudeyras *et al.*, 2008).

Despite the absence of *eefRABCD* in phylogroups A, B1 and C, specific fragments of *eef* coding sequences had been retained suggesting the loss of *eef* has occurred at a common point for the three phylogroups, that being the diversification of phylogroups A, B1 and C from phylogroups B2, D E, F and G. As mentioned different genetic profiles and metabolic capabilities are seen between the phylogroups of *E. coli* species. Thus, it is possible that in phylogroups A, B1 and C, this loss of *eef* occurred

at this time point in which *E. coli* was splitting between lineages with wide metabolic capabilities and those with colonisation/virulence associated lifestyles. Those associated with more commensal roles may not require the efflux of whatever the natural substrate of *eef* is and therefore a reduction in selection pressure has resulted in the loss of the operon. Alternatively a genome event such a deletion may have occurred at the diversification of A, B1 and C from B2, D, E, F and G and the loss of the operon had no significant growth defect, hence the expansion of the A, B1 and C phylogroups despite the lack of this operon. Another alternative is that possession and expression of the *eef* operon imposes some fitness cost if it is not needed.

However, this phylogenetic divide does not explain the lack of operon in ST69 and ST182. The loss of *eef* in these STs suggests the operon has been lost multiple times as *eefRABCD* was generally highly conserved across phylogroups D and E yet ST69 and ST182 had lost additional regions of the CDS when compared to phylogroups A, B1 and C. Both of these STs are known to cause human infections as ST69 is a well-known UPEC lineage and ST182 is associated with foodborne ETEC infections. The loss of this system is potentially due to the specific habit and lifestyle of these *E. coli* phylogroups. Perhaps the main role of *eef* is in fact not host colonisation but instead cell signalling or biofilm production, and that the loss of the operon is not significantly detrimental to the strain survival in their respective niches.

Finally, in addition *to E.* coli and *Klebsiella spp.* the *eef* operon was present in *E. marmoteae, E. albertii,* and *Shigella spp.,* though predominantly *S. dysenteriae. E. coli* and *Shigella* are highly similar genera and could be regrouped as a single genus, however reclassification of *Shigella* has not occurred due to its clinical significance (van den Beld and Reubsaet, 2012, Brenner *et al.*, 1972, Pettengill *et al.*, 2015, Sahl

et al., 2015). In *S. dysenteriae* the first two genes of the *eef* operon, *eefR* and *eefA,* which encode the TetR regulator and PAP were either partial sequences or completely absent (Table 6.9). Both Prokka and ISEscan identified markers of ISs in the region of *eefR* and *eefA.* IS elements are small transposable elements that require integration into plasmids or bacteriophages for their transferal between cells. The integration of IS elements into the genome is associated with early stage genome degradation, and high numbers are seen in the genomes of bacteria that have recently become host-specific (Moran and Plague, 2004). Genomes of *Shigella* spp. are known to have a higher number of IS elements than other members of the Gammaproteobacteria due to their recent host-specific adaptations (Hawkey *et al.*, 2020, Yang *et al.*, 2005). The median number of IS elements in gammaproteobacteria is 24, yet *Shigella* spp. have been found to have over 300 copies (Touchon and Rocha, 2007). Therefore, it is possible that degradation and loss of the *eef* operon is occurring in *S. dysenteriae* as it is adapting and becoming host-specific.

6.14 Key findings

- *E. coli* possess a seventh and previously uncharacterised RND pump, EefABC.
- EefABC is part of a five gene operon that encodes its own TetR regulator and an MFS pump.
- *eefABC* is highly conserved across phylogroups B2, D, E and F but has been lost from A, B1, and C.
- Those that have lost the operon have retained highly conserved fragments of the *eef* CDS.
- The EefABC pump does not transport common classes of antibiotics.

- The operon co-occurs with genes involved in iron metabolism.
- The *eef* operon is present in additional *Escherichia* spp. and a partial operon is present in *S. dysenteriae* however it appears to be in the early stages of degradation.

6.15 Future work

- In the first instance, create additional strains with further gene inactivations and repeat MIC experiments. Creation of *E. coli* ATCC 25922 Δ*acrB* Δ*acrF* Δ*eefB and E. coli* ATCC 25922 Δ*acrB* Δ*acrF* Δ*eefABC* backgrounds. Furthermore transform K-12 Δ*acrB* Δ*acrF* with the pET21a *eefABC* and pET24a constructs. Further inactivation of *acrB* and *acrF* may enable subtle changes in MICs to be detected.
- To determine any possible substrates of EefABC*,* the Biolog phenotype microarray screening platform would allow the testing of hundreds of compounds. This could aid the detection of a substrate and possible function for the RND pump.
- Determine the effects of *eefABC* on the virulence of *E. coli* using the *Galleria mellonella* model in the first instance. Gain of function experiments could determine if the presence of *eefABC* could increase the virulence of non-virulent strains such as K-12, while the effect of *eefABC* inactivation on virulence could also be studied using the same virulence model.
- Measure expression of the *eefRABCD* operon under laboratory conditions and confirm *eef* is expressed in *E. coli* unlike in *K. aerogenes.* Once it has been confirmed *eef* is expressed, it would be possible to determine whether

eefRABCD is involved in the tolerance of extreme pH, low nutrient or anaerobic conditions using growth kinetics experiments and reporter assays.

- Construction of reporter constructs such as a promoter-gfp fusion. This would allow the quantification of activity at the promoter of *eefRABCD* in each condition.
- Co-express *eefABC* and *eefD* in the same cell. Determine whether all four genes are needed to have a detectable phenotype changes.
- Study the effect of *eefABC* inactivation and gain of function on additional phenotypes such as biofilm production and motility.
- Determine whether there is a fitness cost to possessing the *eef* operon. Generate strains that lack the regulator *eefR, eefABCD,* and *eefRABCD.*
- Investigate the high conservation of *eef* remnants in strains that have lost the operon. Measure transcriptional activity of the ORF present in strains that lack *eefRABCD, ycjD,* using RT-PCR to determine whether this is in fact a functional gene.
- Due to the high conservation of the *eef* remnants present in strains lack the operon, deletion of the *ycjD* region and subsequent measurement of phenotypic impact would help elucidate whether this region is of importance.
- Assess open access TSS and RNA data to determine whether there are any small RNAs near *fabI* and *sapF* which might explain the high conservation of *ycjD*.
- Elucidate the role of EefR and assess whether it can directly regulate any additional RND operons.
- Determine the protein structure of EefABC.

Chapter 7 Final Discussion

RND efflux pumps are highly important to the lifestyle of Gram-negative bacteria though they are best characterised due to their role in AMR and drug efflux. The findings of this study confirmed high levels of conservation are seen across RND efflux pumps, especially AcrABZ-TolC which is known to export a wide range of antimicrobial and biocide compounds. High levels of conservation across RND systems that are not predominantly associated with antimicrobial efflux such as AcrD and MdtEF was also detected, supporting the notion that the main role of efflux pumps is not for the expulsion of antimicrobials but instead these systems have more fundamental roles in bacteria which are currently poorly understood. Furthermore, the lack of protein diversity demonstrated in Chapter 3 implies these systems are broadly under a high selection pressure supporting a critical biological role.

Much of our understanding of RND structure and function has arisen from studies using *E. coli* K-12 and K-12 derivatives yet highly conserved differences were seen between the sequences of K-12 and in particular, phylogroups B2, D and F. Though K-12 was originally isolated from a patient sample, the substrain MG1655 used in this study and frequently used by others, has been through rounds of mutagenesis and is well known to have a frame shift resulting in impaired pyrimidine biosynthesis (Jensen, 1993, Hobman *et al.*, 2007). As a result, growth of K-12 in nutrient limited conditions is significantly impaired without uracil supplementation. Furthermore, the K-12 genome is known to be smaller than those of pathogenic *E. coli* and therefore cannot be used a fair representative of infection causing *E. coli* (Horesh *et al.*, 2021, Hayashi *et al.*, 2001). Moreover, sequence variation was seen between the *E. coli* K-12 MG1655 protein sequences and other sequences of ST10 to which K-12 belongs

suggesting it is not even a good representative of the ST to which it belongs. The use of single model organisms for a highly diverse species is therefore impractical and limiting.

While no clear phenotype could be linked to the AcrA T104A, AcrB H596N and TolC A233T substitutions, the high conservation of these across ExPEC associated lineages of *E. coli* suggests slight structural differences between AcrABZ-TolC in K-12 and that of MDR and virulent strains of *E. coli.* As current structural work and EI studies are performed in laboratory strains of Gram-negative bacteria such as *E. coli* K-12 and *S.* Typhimurium SL1344, it raises questions about the efficacy of using these model strains to design compounds for use in the wider clinical setting because while these strains may have once been pathogens, in many cases they have mutated considerably over decades of passage and storage, in some cases losing their ability to cause disease. Furthermore, these strains represent a tiny fraction of the true sequence diversity present in the environment.

The link between efflux conservation and phylogenetic relationship was seen in both Chapter 5 and Chapter 6, suggesting efflux has been highly significant in the evolution of *E. coli.* The absence of *eefRABCD* from K-12 is likely why it has gone unreported for so long. This operon was found to be highly conserved across phylogroups B2, D, E and F, yet lost from A, B1 and C. Due to the proximity of the *eef* operon to the promoter site of *fabI,* it is highly likely that this efflux pump is transcribed at some basal level under laboratory conditions, though deletion of the efflux pump system had no effect on drug susceptibility, nor did deletion of the associated MFS pump. Once again this supports the hypothesis that the main role of RND pumps is not the transport of antimicrobials. The only potential substrates of *eefABC* identified

where the dyes ethidium bromide and rhodamine 6G. While the high conservation of the operon implies an important role in bacterial survival, the operon has been lost in A, B1, and C phylogroups, and lost in separate events in ST69 (D) and ST182 (E), though CDS regions of *eef* have been retained in all lineages that lack the operon and are highly conserved. The remnants of *eef* have been annotated as *ycjD,* and ORF that runs in the opposite direction the *eefRABCD.* This is likely a coincidence, though could also be due to the location of the *fabI* promoter or an unknown regulatory feature.

Fundamentally our understanding of RND efflux pumps is limited and the use of a handful of strains to characterise their functions is detrimental to enhancing our understanding of the wider role of these systems. Through the analysis of thousands of *E. coli* assemblies it is clear that not all *E. coli* strains have six RND pumps; some have six while others have seven. Furthermore, individual STs have independent evolutionary patterns as ST11 lacks a functional AcrF pump and both ST69 and ST182 lack the *eefRABCD* operon. Therefore, future approaches to design new therapeutics and EIs that specifically target RND efflux pumps and their associated PAPs and OMPs must access sequence conservation across a wider selection of lineages and pathovars to determine; 1) whether the target is always present 2) the residue or residues required for compound interaction are widely conserved.

Chapter 8 Appendices

8.1 Sequences of RND genes used in Chapters 3, 4 and 5

Gene	Length	Gene sequence
acrA	1194	>NC_000913.3:484426-485619 Escherichia coli str. K-12 substr. MG1655, complete genome TTAAGACTTGGACTGTTCAGGCTGAGCACCGCTTGCGGCTTGCTGGTTATTAT CAGCGGTAACTTCTTGTGCTTTTACCTGGACACCAGGACGCACTTTCTGCAGC CCACTTATTACTACGCGATCGCCTGCTTTCAGACCTTCTGTCACCAGCCACTT ATCGCCAATAGCCTGGCTTGCAACGATCGGACGGGTTTCCACTTTGTCATCCG CGCCAACTACCAGTACGGTGGCATCGCCACGCGGCGTACGGGTTACGCCCTGT TGCGGGACTAAAATAGCGTTTGGATTAAGCCCTTCTTCCAGACGTGCGCGCAC GAACATACCCGGCAGCAGAGTGTGATCCGGGTTCGGGAAGATAGCGCGTAGGG TGATAGACCCAGTGGTCTGATCAACGGTAACGTCAGAGAATTCCAGCGTACCG TCCTGCGGGAACTTAATGCCGTCACTGGTGATCAGTGACACTTTGGCTTTGCC GTTCTCTTGTTTCAGCGTGCCATTCGCCAGTTCCTGTTTCAGGCGCAGGAAGT CGTTGCTGGACTGGGTCACATCAACGTAGATCGGATCAAGTTGCTGCACGGTT GCCAGCGCAGTCGCCTGACCGTTCTGTACCAATGCGCCTTCCGTCACGTTCGA CTTACCAATGCGACCGCTAATCGGAGAGGTGACTTTGGTGTAAGCCAGATTGA TCCGCGCAGTTTCAACGGCAGCTTTCGCCGCAGTTACCGCAGCATTCGCCTGT TGCGCATCAGCCAGAGCCTGATCGTACTCTTGCTTACTGATGTACTGAGTACC GAGCAGTTTCTGATAACGATTCACCGTCAATTGCGCGATATTGGCTGCAGCCT GGGCTTTCGCCAGATCACCTTTCGCACTGTCGTATGTCGCCTGATAGGTCGCA GGATCAATCTGATAGAGAGAGACACCTGCTTCGATGTCGCTACCTTCTTTGAA ATTACGCTTCAGGATAATCCCGCTAACTTGAGGACGAACTTCTGCGATCCGGT AGGCACTGGTGCGACCCGGAAGCTCGGTTGTGATCTGCAGAGGTTCAGTTTTG ACTGTTACTACGCCAACGGCGGGCATCTGCTGGCCACCTTGTTGGGCCTGTTT GTCGTCACATCCTGTTAGGGCTAAGCTGCCTGAGAGCATCAGAACGACCGCCA GAGGCGTAAACCCTCTGTTTTTGTTCAT
acrB	3150	>NC_000913.3:481254-484403 Escherichia coli str. K-12 substr. MG1655, complete genome TCAATGATGATCGACAGTATGGCTGTGCTCGATATCTTCATTCTTGCGGCTAA AGCGGCGGCGAACCACCACAAAGAATACCGGAACGAAGAAGATTGCCAGTACC GTTGCGGTCACCATCCCGCCCATTACACCGGTACCTACTGCGTTCTGCGCGCC GGAACCAGCACCAGTACTGATAACCAGCGGCATAACGCCGAGGATAAACGCCA GCGAGGTCATCAGGATCGGACGTAAACGCATCCGCACCGCATCAAGCGTCGCT TCAATCAGACCTTTACCTTCTTTATCCATCAAGTCTTTGGCGAATTCGACGAT AAGGATCGCGTTCTTCGCCGACAACCCAATGGTTGTGAGCAGGCCTACCTGGA AGTAAACGTCATTGGTCAGGCCACGGAAGGTGGCAGCCAGCAACGCACCGATA ACCCCCAGCGGAACGACCAGCATAACGGAGAACGGAATCGACCAGCTCTCGTA CAGCGCCGCCAGACACAGGAACACGACAATCAACGAAATCGCGTACAGTGAAG GTGCCTGGTTGCCGGAGAGACGTTCCTGATAGGACATCCCCGTCCAGTCATAG CCAACACCGGTAGGCAGTTTGCTCGCCAGTTGTTCCATCAGCTCCATTGCTTC ACCGGTACTTTTACCCGGTGCCGCCTGGCCTAAGATTTCCATGGATGGCAGGC CGTTGTAACGTTCCAGACGCGGCGAACCGTACTCCCAACGAGAAGAGGAGAAC GCCGAGAATGGCACCATCTGACCATCAGCAGCACGAACATACCAGTCGCCGAT ATCATCCGGCAGCATACGGTATTTCGCTTCTGACATGACATAAACTTTCTTCA CACGACCGCGGTCGATAAAGTCGTTCACATAGCTGCCGCCCCATGCAGCGCCC AGAGTGGTGTTAATGTCGTTGATAGAAACACCCAGCGCCTGCGCTTTTTCCTG GTCGATATCAATCTTAAACTGCGGGGTATCTTCCAGACCGTTTGGACGTACGC TGGTCAACATATCAGGGTGCTTCGCTGCTTCTGCAAGCAACTGGTTACGCGCC TGAGTCAGTTTTTCGTGACCAAGGCCAGCCTGGTCAATCAGCTCAAAGTCAAA GCCGGTTGCAGTACCCAGTTCCACGATTGCGGGCAGGTTAAAGGCGAAAACCA

		TCGCATCTTTGATTTGCGAGAAAGCGCGTGTTGCACGCATGGTAATCGCTTCA ACTTTGTTTTCTTCGCCCGGACGATCGGCCCAGTCCTTCAAGGAAACGAACGC AATACCGGTATTCTGACCACGTCCCGCAAAGCCGAAGCCGTTAACGGCGAACA CCGACTCAACGTTGTTCTTTTCTTTGGTCAGATAGTAATGCGTTACCTCATTG AGCACTTTCTGTGTACGTTCCTGCGTTGCACCTGCTGGCAGCTGAACCATGGT CATAAACACGCCCTGGTCCTCATCTGGCAAGAAGGAGCTTGGCAGACGCACGA ACAGATAGGCCATGCCGACCACGATGATCAGATACAGCACCAGGTAACGCCCC GTACTGCGCAGAATACCGCCTACGCTGTCGGTGTAGTGGTGCGTGCTCTTCTC GAACATGCGGTTAAACCAGCCGAAGAAGCCTTTTTTACCTTCCCCGTGATCGC CTTTGGCAATCGGTTTCAGCATGGTGGCACAAAGAGCTGGAGTCAGGATCAAC GCCACCAGTACCGACAGCGCCATTGCTGAAACAATGGTAATAGAGAACTGACG ATAGATAGCACCAGTAGAACCGCCAAAGAAGGCCATCGGTACGAATACCGCCG ACAGTACCATCGCGATACCGACCAGAGCGCCCTGAATCTGCCCCATCGACTTA CGGGTAGCTTCTTTTGGCGGCAAACCTTCTTCCGCCATAACACGCTCAACGTT TTCTACCACAACGATGGCGTCATCCACCAACAGGCCGATGGCGAGCACCATCC CGAACATTGTTAGCGTGTTTATCGAGAAGCCAAAGGCGGCAAGGACGGCAAAG GTCCCGAGCAATACCACCGGTACGGCAATGGTCGGAATCAACGTCGCGCGGAA GTTCTGCAGGAACAGATACATAACCAGGAACACGAGGATGATCGCTTCGACCA GCGTTTTAACCACTTCGTGAATAGAGATTTTCACGAACGGCGTGGTGTCGTAT GGGTAAACAATTTTCAGACCCGACGGGAAGAACGGTTCCATCTTCGCCAGTTC AGCACGGATTGCCGCAGCGGTATCCAGCGCGTTTGCACCGGTCGCCAGCTTGA TCCCCAGACCGGAAGCCGGTTGGCCGTTAAACTCTGCGATGATGTCGTAGTTC TCACCACCCAGCTCAATCTTCGCGACGTCACGCAGCAGCACGCGGGAACCATC CTGATTCACTTTCAGCAGGATTTTGCCGAACTCTTCAGTAGAGGTCAGACGCG TCTGAGCAATAATAGAGGCGTTAAGCTGTTGGCCTTTCACCGGCGGCGTACCA CCGAGCTGACCCGCCGCAACCTGGGCGTTCTGCGCTTTGATGGCGGTAATGAC ATCAACCGGCGTTAGCTGGAATTTGTTCAGCTCATTCGGGTTCATCCAGATAC GCATCGCGTACTGTGAACCGAACAACTGAACATCACCCACGCCCGACGTACGG CTGATGGCATCTTTCATATTCGCCGCCACGTAGTCGGAGATATCCTCCTGCGT CATGGTGCCATCGGTGTTGATAACGCCGACAACCATCAGGAAGCTGCTGGATG ATTTCTCAACGCTCACCCCTTGCTGCTGAACTTCTTGCGGCAGCAACGGCATC GCCAGCTGCAGTTTGTTCTGCACCTGAACCTGCGCGATATCCGCATCAGTACC AGACTCAAAGGTCAGGGTGATCTGCACGGTACCCGTGGAGTCACTGTTAGAGG ACATGTACATCAGGTTATCGATACCGTTCATATTCTGTTCGATAACCTGTGTC ACCGTGTCCTGCACTGTTTTCGCATCAGCGCCGGGGTAGGAGGCGGAGATCGT TACTGCCGGCGGTGCAATCGTAGGATATTGCGCCACCGGCAGTTTGAGGATCG CCAGCCCCCCTGCCAACATGATGATAATGGCGATCACCCACGCAAAAATCGGG CGATCGATAAAGAAATTAGGCAT
acrD	3114	>NC_000913.3:2587595-2590708 Escherichia coli str. K-12 substr. MG1655, complete genome ATGGCGAATTTCTTTATTGATCGCCCCATTTTTGCCTGGGTGCTGGCAATCCT GTTGTGTCTGACAGGTACCCTGGCGATTTTTTCATTGCCCGTTGAACAATACC CCGATCTCGCGCCACCGAATGTGCGAGTGACCGCTAACTATCCCGGCGCATCG GCCCAGACGCTGGAAAACACCGTGACCCAGGTTATCGAGCAAAATATGACCGG CCTCGATAATCTCATGTATATGTCATCTCAGAGCAGTGGCACCGGTCAGGCAT CTGTCACTTTAAGTTTTAAAGCAGGCACCGATCCGGACGAAGCCGTGCAGCAA GTACAAAACCAGCTGCAATCAGCCATGCGAAAGTTACCGCAGGCGGTGCAAAA TCAGGGCGTGACGGTGCGTAAAACCGGCGATACCAACATTCTGACCATTGCCT TCGTCTCTACCGATGGTTCGATGGATAAACAGGATATTGCTGATTATGTTGCC AGTAATATTCAGGACCCGTTAAGCCGCGTGAATGGCGTCGGGGATATCGATGC CTATGGTTCGCAATATTCCATGCGTATCTGGCTGGACCCGGCGAAACTCAACA GTTTCCAGATGACGGCTAAAGATGTCACTGATGCCATTGAGTCACAGAACGCG CAGATTGCGGTTGGGCAACTTGGTGGTACACCTTCCGTCGATAAGCAGGCGCT CAACGCCACCATTAACGCCCAGTCACTGCTGCAAACACCAGAACAGTTCCGCG ATATCACCTTGCGGGTCAATCAGGACGGCTCAGAGGTAAGGCTGGGCGATGTC GCCACCGTCGAAATGGGGGCGGAGAAATACGATTATCTTAGCCGCTTCAATGG

		TAAGCCAGCCTCCGGGCTGGGGGTAAAACTGGCCTCCGGCGCTAACGAAATGG CGACAGCGGAGCTGGTGCTCAATCGTCTCGACGAGCTGGCGCAGTATTTCCCG CATGGACTGGAATACAAGGTGGCGTATGAAACCACCTCGTTTGTTAAAGCCTC CATTGAAGACGTGGTGAAAACGCTGCTGGAAGCTATCGCTCTGGTTTTCCTCG TTATGTATCTGTTCCTGCAAAACTTCCGCGCCACGCTGATACCCACTATCGCC GTGCCGGTGGTGTTGATGGGAACCTTCTCCGTACTTTACGCCTTCGGTTACAG CGTCAACACCTTAACCATGTTCGCGATGGTGCTGGCGATCGGTCTGCTGGTGG ATGACGCCATCGTGGTGGTGGAAAACGTCGAACGTATTATGAGTGAGGAAGGA CTCACTCCTCGCGAAGCCACACGTAAATCGATGGGGCAGATCCAGGGGGCACT GGTCGGGATTGCGATGGTTCTTTCGGCGGTATTTGTACCAATGGCCTTCTTCG GCGGCACCACCGGTGCCATCTATCGCCAGTTCTCTATTACCATTGTTGCGGCG ATGGTGCTGTCAGTACTGGTAGCGATGATCCTCACTCCGGCTCTGTGTGCCAC ACTACTTAAGCCACTGAAAAAAGGTGAGCATCATGGGCAAAAAGGCTTTTTTG CCTGGTTTAACCAGATGTTTAACCGCAACGCCGAACGCTACGAAAAAGGGGTG GCGAAAATTCTCCACCGTAGCCTGCGCTGGATTGTGATTTATGTCCTGCTGCT TGGCGGCATGGTGTTCCTGTTCCTGCGTTTGCCGACGTCGTTCTTACCGCTGG AAGACCGTGGCATGTTTACTACCTCGGTACAGTTGCCCAGCGGTTCAACGCAA CAACAGACCCTGAAAGTCGTTGAGCAAATCGAGAAATACTACTTCACCCATGA AAAAGACAACATCATGTCGGTGTTTGCCACCGTTGGTTCTGGCCCTGGGGGTA ACGGGCAAAACGTGGCGCGAATGTTTATCCGCCTGAAAGACTGGAGCGAACGC GACAGTAAGACCGGCACCTCGTTTGCCATTATCGAGCGTGCAACGAAGGCGTT TAACCAAATTAAAGAAGCTCGCGTTATCGCCAGCAGCCCGCCAGCAATTAGCG GTCTTGGTAGTTCTGCAGGTTTTGATATGGAGTTGCAGGACCACGCTGGAGCG GGTCACGATGCGCTGATGGCAGCACGTAATCAGTTGCTGGCGCTGGCGGCGGA AAACCCGGAGCTAACCCGTGTGCGCCATAACGGCCTCGACGACAGTCCGCAGT TGCAGATTGATATCGACCAGCGTAAAGCTCAGGCGCTGGGCGTTGCTATCGAC GATATTAACGACACACTGCAAACCGCCTGGGGTTCGAGCTATGTGAATGACTT TATGGATCGCGGTCGCGTGAAGAAAGTCTATGTGCAGGCAGCTGCGCCGTATC GCATGCTGCCAGATGACATCAATCTCTGGTATGTCCGAAATAAAGATGGCGGC ATGGTGCCCTTCTCTGCTTTCGCGACCTCACGCTGGGAAACAGGCTCGCCGCG TCTGGAACGCTATAACGGTTATTCTGCGGTTGAGATTGTTGGGGAAGCCGCAC CGGGGGTCAGTACCGGTACGGCGATGGATATTATGGAATCGTTAGTGAAGCAG CTGCCAAACGGCTTTGGTCTGGAGTGGACGGCGATGTCGTATCAGGAGCGGCT TTCCGGCGCGCAGGCTCCGGCGCTGTACGCCATTTCCTTGCTGGTGGTATTCC TGTGTCTGGCTGCGTTGTATGAAAGCTGGTCGGTGCCGTTCTCGGTAATGCTG GTCGTGCCGCTGGGGGTAATCGGCGCGCTGCTGGCAACCTGGATGCGCGGGCT GGAAAACGACGTTTACTTCCAGGTGGGCCTGTTAACGGTCATTGGTTTATCGG CGAAAAACGCCATCCTGATCGTCGAGTTTGCTAACGAGATGAACCAAAAAGGC CACGACCTGTTTGAAGCGACGCTCCACGCCTGCCGTCAGCGTTTACGCCCGAT TCTGATGACCTCGCTGGCATTTATCTTCGGCGTATTGCCAATGGCAACCAGCA CGGGTGCCGGTTCCGGTGGTCAGCATGCGGTGGGTACTGGCGTAATGGGCGGG ATGATTTCGGCCACTATTCTGGCTATTTACTTCGTGCCGCTGTTCTTTGTGCT GGTGCGCCGCCGCTTCCCGCTGAAGCCGCGCCCGGAATAA
acrE	1158	>NC_000913.3:3413864-3415021 Escherichia coli str. K-12 substr. MG1655, complete genome ATGACGAAACATGCCAGGTTTTTCCTCCTGCCCTCCTTTATTCTGATCTCCGC GGCTTTAATCGCCGGTTGTAACGATAAGGGAGAAGAGAAAGCTCACGTCGGTG AACCGCAGGTTACCGTTCATATTGTAAAAACGGCCCCGTTAGAAGTTAAGACT GAATTACCAGGCCGCACCAATGCTTATCGTATAGCCGAAGTTCGCCCACAGGT TAGCGGGATCGTACTGAATCGCAATTTCACTGAAGGCAGCGATGTGCAAGCAG GCCAGTCCCTGTACCAGATCGATCCCGCGACCTATCAGGCAAATTATGACAGC GCGAAAGGCGAACTGGCGAAAAGTGAAGCCGCCGCCGCCATCGCGCATTTGAC GGTAAAACGTTACGTTCCGCTCGTGGGTACGAAATACATCAGCCAGCAGGAGT ACGACCAGGCCATTGCTGATGCTCGTCAGGCCGATGCCGCCGTGATTGCCGCA AAAGCCACAGTCGAAAGCGCTCGCATCAATCTTGCTTATACCAAAGTCACTGC GCCAATTAGCGGACGTATCGGCAAATCGACTGTGACCGAAGGCGCTCTTGTCA

		CTAATGGGCAAACGACTGAACTGGCGACTGTCCAGCAGCTCGATCCTATCTAC GTTGATGTGACCCAATCCAGCAACGATTTTATGAGGCTGAAGCAATCCGTAGA GCAAGGAAATTTGCATAAGGAAAACGCCACCAGCAACGTAGAGTTGGTCATGG AAAACGGTCAAACCTATCCCCTGAAAGGTACGCTGCAATTCTCCGATGTGACC GTTGATGAAAGCACCGGCTCCATAACCCTACGTGCTGTCTTCCCTAACCCGCA ACATACGCTTTTGCCGGGTATGTTTGTGCGTGCACGGATTGATGAAGGCGTCC AACCTGACGCCATTCTTATCCCGCAACAAGGCGTTAGCCGCACACCGCGTGGT GATGCAACCGTGCTGATTGTTAACGATAAAAGTCAGGTTGAAGCGCGCCCTGT CGTTGCCAGTCAGGCGATTGGCGATAAATGGTTGATTAGTGAAGGACTGAAAT CTGGCGATCAAGTCATTGTCAGCGGCCTGCAAAAAGCGCGTCCGGGAGAGCAG GTTAAAGCCACTACCGATACCCCCGCAGATACTGCATCGAAGTAA
acrF	3105	>NC_000913.3:3415033-3418137 Escherichia coli str. K-12 substr. MG1655, complete genome ATGGCAAACTTTTTTATTCGACGACCGATATTTGCATGGGTGCTGGCCATTAT TCTGATGATGGCGGGCGCACTGGCGATCCTACAATTGCCCGTCGCTCAGTATC CAACAATTGCACCGCCTGCGGTTTCTGTTTCAGCAAACTATCCGGGCGCTGAT GCGCAGACCGTGCAGGATACGGTGACGCAGGTTATCGAACAGAATATGAACGG TATCGATAACCTGATGTATATGTCCTCCACCAGCGATTCCGCCGGTAGCGTGA CAATTACCCTTACCTTCCAGTCCGGGACCGATCCTGATATCGCGCAAGTGCAG GTGCAGAACAAACTCCAGCTCGCCACGCCGTTGCTGCCGCAGGAGGTTCAGCA GCAGGGGATCAGTGTTGAAAAGTCCAGTAGCAGCTATTTGATGGTGGCGGGCT TTGTCTCTGATAACCCAGGCACCACACAGGACGATATCTCGGACTATGTGGCC TCTAACGTTAAAGATACGCTTAGCCGTCTGAATGGCGTCGGTGACGTACAGCT TTTCGGCGCACAGTATGCGATGCGTATCTGGCTGGATGCCGATCTGCTAAACA AATATAAACTGACACCGGTTGATGTGATTAACCAGTTGAAGGTACAGAACGAT CAGATCGCTGCCGGACAGTTGGGCGGAACGCCAGCGTTACCAGGGCAACAATT GAACGCCTCGATTATTGCTCAGACGCGGTTTAAAAATCCGGAAGAATTCGGCA AAGTGACCCTGCGCGTAAACAGTGACGGCTCGGTGGTACGCCTGAAAGATGTC GCACGGGTTGAACTTGGCGGTGAAAACTATAACGTTATCGCTCGTATCAACGG AAAACCGGCGGCGGGCCTGGGGATTAAGCTGGCAACCGGCGCGAATGCTCTCG ATACCGCGAAAGCCATTAAGGCAAAACTGGCGGAATTACAGCCATTCTTCCCG CAGGGAATGAAGGTTCTCTACCCTTATGACACCACGCCATTCGTCCAGCTTTC TATTCACGAAGTGGTAAAAACGCTGTTCGAAGCCATTATGCTGGTGTTCCTGG TGATGTATCTGTTCTTGCAGAATATGCGAGCAACGCTGATCCCCACCATTGCG GTACCCGTGGTGTTGTTAGGGACGTTTGCCATCCTCGCCGCTTTTGGTTACTC CATCAACACACTAACGATGTTCGGGATGGTGCTTGCCATCGGGCTGCTCGTCG ATGATGCGATAGTGGTGGTGGAGAACGTCGAGCGCGTGATGATGGAGGATAAG CTCCCGCCAAAAGAAGCGACGGAAAAATCGATGTCGCAAATTCAGGGCGCACT GGTGGGTATCGCGATGGTGCTGTCAGCGGTATTTATTCCGATGGCATTCTTCG GCGGTTCTACTGGGGCAATTTATCGCCAGTTCTCTATCACCATCGTTTCGGCA ATGGCGCTTTCTGTTCTGGTGGCATTGATTCTTACCCCTGCGTTATGTGCAAC GCTGCTTAAACCCGTCTCTGCTGAGCATCACGAAAATAAGGGCGGTTTCTTCG GTTGGTTTAATACCACCTTCGATCATAGCGTTAACCACTACACCAACAGCGTC GGCAAAATCCTCGGATCCACAGGACGATATTTACTGATCTATGCGCTGATTGT TGCAGGAATGGTGGTGTTGTTTTTACGTCTTCCGTCTTCCTTCTTACCTGAAG AGGATCAGGGTGTCTTTCTGACCATGATTCAGTTACCCGCTGGCGCGACGCAA GAGCGGACGCAAAAAGTGTTGGATCAAGTTACGGATTACTATCTGAAGAACGA GAAAGCGAACGTTGAAAGTGTCTTTACGGTTAACGGCTTTAGCTTCAGCGGCC AGGCACAAAACGCCGGTATGGCCTTCGTCAGTCTGAAACCGTGGGAAGAGCGT AATGGTGACGAAAACAGTGCGGAAGCGGTAATCCATCGTGCCAAAATGGAATT GGGCAAGATCCGCGACGGTTTTGTCATTCCATTCAATATGCCAGCCATTGTTG AACTGGGCACGGCAACGGGTTTCGACTTTGAGTTAATTGATCAGGCTGGGCTG GGTCACGATGCCCTAACCCAGGCCCGTAACCAGTTGCTTGGTATGGCGGCGCA ACATCCTGCCAGCTTAGTCAGCGTGCGCCCTAATGGCCTGGAAGACACCGCGC AGTTTAAACTGGAAGTTGACCAGGAAAAGGCGCAGGCATTAGGTGTTTCACTT TCTGACATCAATCAGACCATTTCAACGGCGCTGGGTGGGACTTACGTTAACGA

		CTTCATCGACCGTGGCCGCGTGAAAAAGTTGTATGTTCAGGCGGATGCCAAAT TCCGTATGCTGCCAGAAGATGTCGATAAACTTTATGTCCGCAGCGCCAACGGC GAAATGGTGCCATTCTCGGCCTTTACCACTTCACATTGGGTGTATGGCTCTCC GCGACTGGAACGCTACAACGGTCTGCCGTCAATGGAGATTCAGGGGGAAGCCG CGCCAGGAACCAGTTCCGGCGATGCCATGGCGTTGATGGAAAACCTTGCGTCA AAATTACCTGCGGGCATTGGTTATGACTGGACGGGTATGTCGTATCAGGAACG CTTATCGGGAAACCAGGCTCCCGCTCTGGTAGCAATTTCCTTTGTGGTTGTTT TCCTGTGCCTTGCTGCACTCTATGAAAGCTGGTCAATTCCTGTCTCGGTTATG TTGGTAGTGCCGTTAGGGATTGTCGGCGTGCTGCTGGCGGCGACACTCTTTAA TCAAAAAAATGACGTCTACTTTATGGTGGGCTTGCTAACGACAATTGGCTTGT CGGCCAAAAACGCTATTTTGATCGTTGAGTTCGCTAAAGATCTCATGGAGAAA GAGGGTAAAGGTGTTGTTGAAGCGACACTGATGGCAGTACGTATGCGTCTGCG TCCTATCCTGATGACCTCTCTCGCCTTTATTCTCGGCGTATTACCGCTAGCTA TCAGTAACGGTGCCGGCAGTGGCGCGCAGAACGCTGTGGGTATCGGGGTAATG GGAGGAATGGTCTCTGCAACGTTGCTGGCAATCTTCTTCGTACCGGTGTTCTT TGTGGTGATCCGCCGTTGCTTTAAAGGATAA
acrZ	150	>NC_000913.3:794773-794922 Escherichia coli str. K-12 substr. MG1655, complete genome ATGTTAGAGTTATTAAAAAGTCTGGTATTCGCCGTAATCATGGTACCTGTCGT GATGGCCATCATCCTGGGTCTGATTTACGGTCTTGGTGAAGTATTCAACATCT TTTCTGGTGTTGGTAAAAAAGACCAGCCCGGACAAAATCATTGA
cusA	3144	>NC_000913.3:598714-601857 Escherichia coli str. K-12 substr. MG1655, complete genome ATGATTGAATGGATTATTCGTCGCTCGGTGGCGAACCGTTTTTTGGTGCTGAT GGGCGCGTTGTTTCTGAGCATCTGGGGCACCTGGACCATCATTAATACGCCAG TGGATGCGCTGCCGGATCTCTCCGATGTGCAGGTTATTATTAAAACCAGCTAT CCCGGTCAGGCACCACAAATCGTTGAAAATCAGGTGACTTATCCGCTAACCAC CACCATGTTGTCGGTGCCTGGCGCGAAGACTGTGCGCGGTTTCTCACAGTTTG GCGACTCTTATGTGTATGTCATTTTCGAAGATGGCACCGATCCGTACTGGGCG CGTTCGCGGGTGCTGGAGTACCTCAACCAGGTACAGGGTAAGCTGCCTGCGGG AGTCAGCGCCGAGCTGGGGCCAGATGCCACGGGTGTTGGCTGGATCTATGAAT ATGCACTGGTGGATCGCAGCGGTAAGCACGATCTGGCCGATTTACGCTCATTA CAGGACTGGTTTCTCAAATATGAGCTGAAAACCATCCCTGACGTTGCGGAAGT GGCGTCGGTGGGCGGTGTGGTGAAAGAGTATCAGGTGGTTATCGATCCCCAGC GCCTGGCGCAGTATGGCATCAGTCTCGCCGAAGTAAAAAGCGCGCTGGATGCT TCAAACCAGGAAGCGGGCGGTTCGTCGATCGAACTGGCGGAAGCGGAATATAT GGTGCGCGCCAGCGGCTATCTGCAAACGCTCGACGACTTTAATCACATCGTTT TAAAAGCCAGTGAAAATGGCGTGCCCGTTTATCTGCGCGATGTTGCGAAGGTC CAGATTGGCCCGGAGATGCGCCGGGGCATTGCCGAACTAAACGGCGAAGGCGA AGTGGCGGGCGGGGTGGTGATCCTGCGATCCGGCAAAAACGCCCGAGAAGTGA TCGCCGCCGTGAAGGACAAACTGGAAACGCTGAAAAGTAGTCTGCCGGAAGGC GTGGAGATAGTTACAACATACGATCGCAGCCAGCTCATTGACCGCGCCATCGA CAACCTCAGCGGCAAGTTGCTGGAAGAGTTTATTGTGGTGGCGGTAGTCTGTG CGCTGTTTCTCTGGCATGTGCGCTCGGCGCTGGTGGCGATTATTTCGTTGCCG CTGGGGTTGTGTATTGCTTTTATTGTCATGCACTTCCAGGGACTGAATGCCAA TATTATGTCGTTGGGTGGCATTGCGATTGCCGTCGGGGCGATGGTCGATGCTG CTATCGTCATGATCGAGAATGCGCATAAACGGCTGGAAGAGTGGCAGCACCAG CATCCTGACGCCACGCTGGATAATAAAACGCGCTGGCAGGTGATCACCGATGC GTCTGTTGAAGTGGGGCCGGCGCTATTTATCAGTCTGCTGATTATCACGTTGT CGTTTATCCCGATCTTCACCCTGGAAGGGCAGGAAGGGCGTCTGTTTGGCCCG TTGGCGTTCACCAAAACGTATGCGATGGCGGGTGCGGCGCTGCTGGCGATCGT AGTGATCCCGATCCTGATGGGCTACTGGATCCGTGGCAAAATTCCGCCGGAAA GCAGTAACCCGCTCAATCGCTTTTTGATTCGTGTTTATCATCCGCTGTTGCTG AAAGTACTGCACTGGCCGAAAACCACGCTGCTGGTGGCGGCGCTTTCGGTGCT GACGGTTCTCTGGCCGCTCAATAAAGTTGGCGGGGAATTTTTACCGCAGATCA ATGAAGGCGACTTGTTGTATATGCCATCGACGCTGCCGGGGATTTCCGCAGCA

		GAGGCGGCGAGTATGCTGCAAAAAACCGACAAGCTAATTATGAGCGTACCTGA AGTGGCGCGGGTATTTGGCAAAACCGGGAAAGCGGAAACCGCCACCGATTCTG CTCCGCTGGAGATGGTAGAAACGACCATCCAGCTTAAGCCGCAGGAGCAGTGG CGGCCAGGCATGACGATGGACAAAATCATTGAGGAACTGGATAACACCGTGCG GCTGCCGGGGCTGGCGAATCTGTGGGTGCCGCCAATTCGTAACCGTATCGATA TGCTCTCAACCGGCATTAAAAGCCCCATCGGCATTAAAGTTTCCGGCACTGTG CTGGCGGATATCGACGCGATGGCTGAGCAAATTGAAGAAGTAGCGCGAACGGT GCCAGGCGTAGCTTCTGCGCTTGCCGAGCGGCTGGAAGGTGGGCGCTATATCA ACGTTGAGATTAACCGTGAAAAAGCCGCACGTTACGGTATGACGGTGGCGGAT GTGCAGTTGTTTGTGACTTCTGCGGTGGGCGGGGCGATGGTTGGCGAAACGGT GGAAGGGATTGCCCGTTATCCAATTAATCTGCGTTATCCGCAAAGCTGGCGCG ATAGTCCGCAGGCACTGCGCCAGCTGCCGATCCTGACGCCGATGAAGCAGCAA ATCACCCTGGCAGACGTGGCCGACATTAAAGTCTCTACCGGACCGTCGATGCT GAAAACCGAGAATGCGCGCCCGACGAGCTGGATTTATATCGATGCCCGCGATC GTGACATGGTGTCGGTGGTTCACGATTTGCAAAAAGCGATAGCTGAAAAAGTG CAGTTAAAACCGGGCACCAGCGTGGCATTCTCCGGGCAGTTCGAGCTACTGGA GCGCGCCAACCATAAGCTTAAACTCATGGTGCCGATGACGTTGATGATTATCT TCGTGCTGTTGTATCTGGCGTTCCGTCGGGTGGGCGAAGCGTTGCTGATTATC AGCAGCGTACCGTTTGCGCTGGTGGGCGGCATCTGGTTGCTGTGGTGGATGGG CTTTCATCTTTCCGTGGCGACGGGCACTGGCTTTATCGCCCTCGCCGGGGTCG CCGCCGAATTTGGCGTGGTGATGCTGATGTATTTACGTCACGCCATAGAGGCC GTGCCGTCGTTGAATAATCCACAAACATTCAGCGAGCAGAAGCTGGATGAGGC GTTATATCACGGCGCGGTCCTGCGCGTGCGCCCGAAAGCGATGACGGTGGCGG TGATTATCGCCGGTCTGCTGCCGATTCTGTGGGGAACGGGGGCTGGTTCAGAG GTGATGAGCCGGATTGCCGCGCCGATGATTGGCGGCATGATCACCGCACCTTT GCTGTCGCTGTTTATTATCCCGGCGGCGTATAAGCTGATGTGGCTGCACCGAC ATCGGGTACGGAAATAA
cusB	1224	>NC_000913.3:597479-598702 Escherichia coli str. K-12 substr. MG1655, complete genome ATGAAAAAAATCGCGCTTATTATCGGCAGCATGATCGCGGGCGGTATTATTTC TGCGGCAGGTTTTACCTGGGTTGCAAAGGCGGAACCGCCTGCAGAAAAAACGT CGACCGCAGAACGTAAAATCTTATTCTGGTACGACCCAATGTATCCCAATACG CGGTTCGATAAACCAGGTAAATCGCCGTTTATGGATATGGATCTGGTGCCGAA ATATGCCGATGAAGAGAGTTCTGCGTCTGGTGTGCGCATTGACCCGACTCAGA CGCAGAATCTGGGGGTGAAAACGGCTACCGTCACGCGCGGACCGCTGACTTTT GCCCAGAGTTTCCCGGCGAATGTCAGTTACAACGAGTATCAGTATGCCATTGT GCAGGCTCGCGCTGCCGGGTTTATCGACAAGGTGTATCCGCTTACCGTGGGCG ATAAAGTACAAAAGGGCACACCGCTTCTCGACCTGACCATTCCTGACTGGGTG GAAGCGCAGAGTGAGTATTTACTGCTGCGCGAAACCGGCGGTACGGCGACCCA GACTGAAGGCATTCTTGAGCGGCTGCGACTGGCGGGAATGCCGGAGGCGGATA TTCGCCGCCTGATCGCCACGCAAAAAATCCAGACTCGCTTTACGCTCAAAGCG CCCATTGATGGCGTGATCACCGCGTTTGATCTGCGCGCGGGAATGAATATCGC CAAAGATAACGTGGTAGCGAAAATTCAGGGTATGGACCCGGTGTGGGTCACTG CTGCGATCCCGGAGTCTATCGCCTGGCTGGTGAAAGATGCCTCGCAGTTTACC CTCACCGTTCCGGCGCGACCGGATAAAACACTCACCATCCGCAAATGGACGCT GCTACCTGGCGTGGATGCCGCGACCCGCACGCTGCAGCTGCGTCTGGAAGTCG ACAACGCCGACGAGGCGCTAAAACCGGGCATGAACGCCTGGTTGCAACTCAAC ACCGCCAGCGAACCGATGCTGCTCATTCCGTCACAAGCGCTGATTGATACCGG CAGCGAACAGCGGGTGATTACCGTTGATGCCGACGGGCGCTTTGTACCGAAAC GCGTTGCTGTTTTCCAGGCATCGCAAGGCGTCACCGCATTACGCTCTGGTCTG GCGGAAGGTGAAAAGGTGGTTTCCAGCGGCCTGTTCCTGATTGATTCCGAAGC CAATATTTCTGGCGCACTGGAGCGGATGCGCTCTGAAAGTGCTACCCATGCGC ATTGA
cusC	1374	>NC_000913.3:595600-596973 Escherichia coli str. K-12 substr. MG1655, complete genome

		ATGTCTCCTTGTAAACTTCTGCCATTTTGTGTGGCCCTTGCGCTAACCGGTTG TTCACTGGCACCGGATTATCAGCGTCCGGCAATGCCCGTGCCGCAGCAGTTCT CACTCAGCCAGAACGGCCTGGTTAACGCAGCAGATAACTATCAGAACGCGGGC TGGCGCACCTTTTTTGTTGATAATCAGGTGAAGACGCTGATTAGCGAGGCGCT GGTGAATAATCGGGATTTGCGCATGGCGACGCTGAAAGTGCAGGAGGCACGGG CGCAATATCGTCTGACCGATGCCGACCGCTACCCACAGCTCAATGGCGAGGGC AGCGGCAGCTGGAGCGGTAATCTTAAAGGGAATACCGCCACGACACGGGAGTT TTCGACCGGCCTTAACGCCAGCTTTGACCTCGATTTTTTCGGTCGCTTAAAGA ACATGAGCGAAGCCGAGCGACAAAATTATTTAGCCACTGAGGAAGCTCAGCGC GCGGTGCATATTCTGCTGGTTTCTAATGTCGCGCAAAGCTATTTCAATCAGCA ACTGGCGTATGCGCAATTGCAAATAGCCGAAGAAACGCTGCGTAATTATCAGC AGTCATATGCGTTTGTCGAAAAACAACTGTTGACAGGTAGCAGTAATGTTCTG GCGCTGGAACAGGCTCGCGGGGTGATAGAAAGTACCCGCAGCGACATCGCTAA ACGTCAGGGGGAACTGGCGCAGGCGAATAATGCATTGCAACTGTTGCTGGGAA GCTACGGCAAGCTGCCGCAAGCGCAGACAGTAAACAGCGACAGCCTGCAAAGC GTTAAATTACCGGCGGGATTGTCGTCGCAAATCTTATTGCAGCGCCCGGATAT TATGGAAGCTGAACACGCGTTAATGGCGGCTAATGCCAATATTGGCGCTGCAC GCGCGGCATTTTTTCCGTCTATAAGCCTGACCAGCGGAATATCGACCGCCAGT AGCGATCTATCGTCATTATTTAACGCCAGCAGCGGGATGTGGAATTTTATACC CAAAATTGAGATCCCCATTTTTAATGCCGGACGCAACCAGGCCAATCTGGATA TCGCCGAAATTCGCCAGCAGCAGTCGGTGGTGAATTATGAACAGAAAATCCAG AACGCCTTTAAAGAAGTGGCAGATGCGCTGGCATTACGTCAAAGCCTGAACGA TCAAATTAGCGCCCAGCAGCGTTATCTGGCATCGCTGCAAATTACTTTGCAAC GGGCGCGGGCATTATATCAGCACGGCGCAGTAAGTTATCTGGAAGTGCTGGAT GCCGAGCGTTCTTTATTTGCAACCCGACAAACTTTACTTGATCTGAATTATGC CCGTCAGGTTAACGAAATTTCTTTGTATACCGCACTTGGTGGCGGTTGA
mdtA	1248	>NC_000913.3:2154016-2155263 Escherichia coli str. K-12 substr. MG1655, complete genome ATGAAAGGTAGTTATAAATCCCGTTGGGTAATCGTAATCGTGGTGGTTATCGC CGCCATCGCCGCATTCTGGTTCTGGCAAGGCCGCAATGACTCCCGGAGTGCAG CCCCAGGGGCGACGAAACAAGCGCAGCAATCGCCAGCGGGTGGTCGACGTGGT ATGCGTTCCGGCCCATTAGCCCCGGTTCAGGCGGCGACCGCCGTAGAACAGGC AGTTCCGCGTTACCTCACCGGGCTTGGCACCATTACCGCCGCTAATACCGTTA CGGTGCGCAGCCGCGTGGACGGCCAACTGATAGCGTTACATTTCCAGGAAGGC CAGCAGGTCAAAGCAGGCGATTTACTGGCAGAAATTGACCCCAGCCAGTTCAA AGTTGCATTAGCACAAGCCCAGGGCCAACTGGCAAAAGATAAAGCCACGCTTG CCAACGCCCGCCGTGACCTGGCGCGTTATCAACAACTGGCAAAAACCAATCTC GTTTCCCGCCAGGAGCTGGATGCCCAACAGGCGCTGGTCAGTGAAACCGAAGG CACCATTAAGGCTGATGAAGCAAGCGTTGCCAGCGCGCAGCTGCAACTCGACT GGAGCCGGATTACCGCACCAGTCGATGGTCGCGTTGGTCTCAAGCAGGTTGAT GTTGGTAACCAAATCTCCAGTGGTGATACCACCGGGATCGTGGTGATCACCCA GACGCATCCTATCGATTTAGTCTTTACCCTGCCGGAAAGCGATATCGCTACCG TAGTGCAGGCGCAGAAAGCCGGAAAACCGCTGGTGGTAGAAGCCTGGGATCGC ACCAACTCGAAGAAATTAAGTGAAGGCACGCTGTTAAGTCTAGATAACCAAAT CGATGCCACTACCGGTACGATTAAAGTGAAAGCACGCTTTAATAATCAGGATG ATGCGCTGTTTCCCAATCAGTTTGTTAACGCGCGCATGTTAGTCGACACCGAA CAAAACGCCGTAGTGATCCCAACAGCCGCCCTGCAAATGGGCAATGAAGGCCA TTTTGTCTGGGTGCTGAATAGCGAAAACAAGGTCAGCAAACATCTGGTGACGC CGGGCATTCAGGACAGTCAGAAAGTGGTGATCCGTGCAGGTATTTCTGCGGGC GATCGCGTGGTGACAGACGGCATTGATCGCCTGACCGAAGGGGCGAAAGTGGA AGTGGTGGAAGCCCAGAGCGCCACTACTCCGGAAGAGAAAGCCACCAGCCGCG AATACGCGAAAAAAGGAGCACGCTCCTGA
mdtB	3123	>NC_000913.3:2155263-2158385 Escherichia coli str. K-12 substr. MG1655, complete genome ATGCAGGTGTTACCCCCGAGCAGCACAGGCGGCCCGTCGCGCCTGTTTATTAT GCGTCCTGTGGCCACCACGCTGCTGATGGTGGCGATCTTACTCGCCGGGATTA

TCGGTTATCGCGCCCTGCCCGTTTCGGCGCTGCCGGAAGTGGACTATCCGACC
ATTCAGGTGGTCACGCTCTACCCAGGTGCCAGCCCGGATGTCATGACCTCTGC
CGTTACCGCGCCGCTAGAACGCCAGTTCGGGCAGATGTCTGGCCTGAAACAGA
TGTCGTCGCAAAGTTCCGGCGGTGCGTCAGTTATCACTTTGCAGTTCCAGCTA
ACATTACCGCTCGATGTCGCCGAGCAGGAAGTGCAGGCCGCGATTAACGCTGC
GACCAACTTGTTGCCGAGCGATCTGCCTAACCCGCCGGTTTACAGCAAAGTGA
ACCCGGCAGATCCGCCGATCATGACGCTCGCCGTCACCTCAACCGCCATGCCG
ATGACGCAAGTGGAAGATATGGTGGAAACCCGCGTCGCGCAGAAAATCTCGCA
GATTTCCGGCGTCGGCCTGGTGACGCTTTCCGGCGGTCAGCGTCCGGCTGTTC
GCGTCAAACTTAACGCTCAGGCGATTGCCGCCCTCGGCCTGACCAGCGAAACC
GTGCGCACCGCCATTACCGGCGCTAACGTTAACTCGGCAAAAGGTAGCCTCGA
CGGCCCTTCCCGTGCGGTCACGCTTTCCGCGAACGACCAGATGCAATCCGCCG
AAGAGTATCGCCAGCTAATCATCGCCTACCAGAACGGCGCGCCAATTCGTCTG
GGCGATGTCGCAACTGTAGAGCAAGGTGCAGAAAACAGCTGGCTCGGCGCGTG
GGCGAACAAAGAACAGGCCATTGTGATGAATGTTCAGCGCCAGCCCGGTGCTA
ACATTATCTCCACCGCCGACAGCATTCGGCAGATGCTGCCACAGCTCACTGAG
AGTCTGCCGAAATCGGTGAAGGTGACAGTGCTTTCCGATCGCACCACCAATAT
CCGCGCATCCGTCGATGATACTCAGTTTGAATTGATGATGGCTATCGCGCTGG
TAGTCATGATTATCTACCTGTTTTTGCGCAATATTCCGGCGACCATCATTCCC
GGTGTTGCTGTACCGCTGTCGTTAATCGGCACTTTCGCGGTTATGGTGTTTCT
CGATTTTTCAATCAATAACCTGACACTGATGGCGTTAACTATCGCCACCGGAT
TCGTGGTCGATGACGCCATCGTGGTGATCGAAAACATTTCCCGCTATATCGAA
AAAGGCGAAAAACCGTTGGCGGCGGCGCTCAAGGGCGCAGGTGAAATCGGCTT
TACCATTATCTCGCTGACCTTCTCACTGATTGCGGTGTTGATCCCACTGCTGT
TTATGGGCGATATCGTCGGGCGACTGTTCCGCGAATTTGCTATTACCCTGGCG
GTAGCGATTTTGATCTCAGCGGTGGTGTCGCTGACCCTGACACCGATGATGTG
CGCGCGGATGCTCAGCCAGGAGTCGTTGCGTAAACAGAACCGCTTCTCCCGTG
CCTCGGAAAAAATGTTCGACAGGATAATCGCCGCCTATGGTCGTGGACTGGCG
AAAGTGCTGAATCATCCGTGGCTGACCTTAAGCGTGGCACTCAGCACGCTGCT
GCTTAGCGTGCTGCTGTGGGTGTTCATTCCGAAAGGTTTCTTCCCGGTACAGG
ACAATGGCATTATTCAGGGCACTTTGCAGGCACCGCAATCCAGCTCCTTTGCC
AATATGGCCCAGCGACAACGCCAGGTCGCGGACGTGATTTTGCAGGATCCGGC
AGTGCAAAGCCTGACCTCATTTGTTGGCGTTGATGGCACTAACCCGTCGCTGA
ACAGTGCACGTTTACAAATCAACCTCAAACCGTTGGATGAACGTGATGATCGG
GTGCAAAAAGTCATCGCCCGTCTGCAAACGGCGGTAGATAAAGTGCCGGGCGT
CGATCTCTTCCTGCAACCAACGCAGGATCTGACTATTGATACTCAGGTCAGCC
GCACCCAGTACCAGTTTACCTTGCAGGCCACGTCACTGGATGCGCTCAGTACC
TGGGTGCCACAGTTGATGGAAAAACTCCAGCAACTGCCACAGCTTTCTGATGT
CTCCAGCGACTGGCAGGACAAAGGGCTGGTGGCGTATGTCAATGTTGATCGCG
ACAGCGCCAGCCGTCTGGGGATCAGCATGGCGGATGTCGATAACGCCCTGTAC
AACGCGTTTGGTCAGCGGCTGATTTCCACTATTTATACTCAGGCCAACCAGTA
TCGCGTGGTGCTGGAGCACAACACCGAAAATACCCCAGGCCTCGCGGCGCTGG
ATACCATTCGCCTGACCAGCAGCGACGGCGGCGTGGTGCCGCTAAGCTCAATT
GCCAAAATTGAGCAGCGTTTTGCGCCGCTCTCCATCAACCATCTGGATCAGTT
CCCGGTAACGACCATCTCCTTTAACGTGCCGGATAACTATTCGCTGGGCGATG
CGGTGCAGGCGATTATGGACACCGAAAAGACGCTGAATCTGCCGGTGGATATC
ACCACGCAGTTCCAGGGCAGCACCCTCGCCTTCCAGTCGGCGCTGGGCAGCAC
TGTCTGGCTGATTGTCGCGGCGGTGGTGGCGATGTATATCGTGCTCGGCATTC
TGTACGAGAGCTTTATTCACCCGATCACCATTCTCTCGACGCTACCCACCGCA
GGGGTTGGCGCACTGCTGGCGTTGCTGATTGCTGGTAGCGAACTGGATGTGAT
TGCGATTATCGGCATTATTTTGCTGATCGGTATCGTGAAGAAGAACGCCATCA
TGATGATCGACTTCGCGCTGGCTGCTGAGCGCGAGCAAGGCATGTCGCCGCGC
GAGGCAATCTACCAGGCTTGTCTGTTGCGTTTTCGTCCGATCCTGATGACCAC
TCTGGCGGCTCTGCTTGGCGCGCTGCCGCTGATGTTGAGTACCGGGGTCGGCG
CGGAACTGCGTCGTCCGTTAGGTATCGGCATGGTCGGCGGTCTGATTGTCAGC

		CAGGTGCTGACGCTGTTTACCACGCCGGTGATTTATTTGCTGTTCGACCGCCT GGCATTGTGGACCAAAAGCCGCTTTGCCCGTCATGAAGAGGAGGCGTAA
mdtC	3078	>NC_000913.3:2158386-2161463 Escherichia coli str. K-12 substr. MG1655, complete genome GTGAAGTTTTTTGCCCTCTTCATTTACCGCCCGGTGGCGACGATTTTACTGTC GGTTGCCATTACCCTGTGCGGCATACTGGGCTTCCGTATGCTGCCGGTCGCCC CGCTGCCGCAGGTCGATTTTCCGGTGATTATCGTCAGCGCCTCGCTGCCCGGT GCGTCACCAGAAACAATGGCGTCTTCCGTTGCCACGCCGCTGGAGCGCTCACT TGGGCGCATTGCCGGAGTCAGTGAAATGACCTCCAGCAGTTCGCTCGGCAGCA CGCGTATTATTTTGCAGTTTGATTTTGACCGGGATATCAACGGCGCAGCGCGT GATGTGCAGGCGGCGATCAACGCTGCACAAAGTTTGCTGCCCAGTGGGATGCC CAGCCGCCCGACCTATCGCAAAGCGAACCCGTCGGATGCGCCAATTATGATCC TCACGCTGACGTCCGATACTTATTCGCAGGGTGAACTGTACGATTTCGCCTCG ACGCAGCTGGCTCCGACGATTTCGCAAATCGACGGTGTTGGTGATGTCGATGT CGGAGGCAGCTCACTGCCCGCCGTACGCGTCGGGCTGAATCCGCAGGCGCTGT TTAATCAGGGCGTGTCGCTGGACGACGTACGCACCGCCGTCAGCAATGCCAAC GTGCGTAAACCGCAGGGCGCGCTGGAAGATGGCACTCACCGCTGGCAGATCCA GACCAATGATGAGCTAAAAACCGCCGCTGAATATCAGCCGTTGATTATTCACT ACAACAACGGCGGCGCGGTTCGTCTGGGCGATGTGGCGACGGTGACCGACTCA GTGCAGGATGTGCGCAACGCCGGGATGACCAACGCCAAACCGGCTATTTTACT GATGATCCGCAAACTGCCGGAAGCCAATATTATCCAGACGGTTGACAGCATCC GGGCAAAATTACCGGAGTTGCAGGAAACCATTCCGGCGGCGATTGATCTGCAA ATTGCCCAGGATCGCTCCCCCACCATTCGCGCCTCGCTGGAAGAAGTCGAGCA AACGCTGATTATCTCGGTGGCGCTGGTGATTCTGGTGGTGTTTTTATTCCTGC GCTCGGGTCGCGCCACTATTATTCCCGCCGTTTCGGTGCCGGTTTCGCTGATT GGTACGTTTGCGGCGATGTACCTGTGCGGATTCAGTCTCAATAACCTTTCGTT AATGGCGCTCACCATCGCTACTGGTTTCGTGGTGGATGACGCCATCGTGGTGC TGGAAAACATTGCACGTCATCTGGAAGCGGGAATGAAACCGTTGCAAGCCGCA CTGCAAGGTACTCGCGAAGTCGGTTTTACGGTGCTGTCGATGAGTCTGTCACT GGTGGCGGTGTTCCTGCCGCTGCTGTTGATGGGCGGATTGCCGGGCCGACTGT TACGCGAATTTGCCGTGACGCTTTCTGTCGCCATTGGTATATCGTTGCTGGTT TCTCTGACATTAACGCCAATGATGTGTGGCTGGATGCTGAAAGCCAGCAAGCC GCGCGAGCAAAAGCGACTGCGTGGTTTTGGTCGCATGTTGGTAGCCCTGCAAC AAGGCTACGGCAAGTCACTAAAATGGGTGCTCAATCATACCCGTCTGGTGGGC GTGGTGCTGCTTGGCACCATTGCGCTGAATATCTGGCTGTATATCTCGATCCC GAAAACCTTCTTCCCGGAGCAGGACACTGGCGTGTTGATGGGCGGGATTCAGG CGGATCAGAGTATTTCGTTTCAGGCGATGCGCGGTAAGTTGCAGGATTTCATG AAAATTATCCGTGACGATCCGGCAGTGGATAATGTCACCGGCTTTACAGGCGG TTCGCGAGTGAACAGCGGGATGATGTTTATCACCCTCAAGCCACGCGACGAAC GCAGCGAAACGGCGCAGCAAATTATCGACCGTCTGCGCGTAAAACTGGCGAAA GAACCGGGGGCGAATCTGTTCCTGATGGCGGTACAGGATATTCGCGTTGGTGG GCGTCAGTCGAACGCCAGCTACCAGTACACGTTGTTATCCGACGACCTGGCGG CACTGCGAGAATGGGAGCCGAAAATCCGCAAAAAACTGGCGACGTTGCCGGAA CTGGCGGACGTGAACTCCGATCAGCAGGATAACGGCGCGGAGATGAATCTGGT TTACGACCGCGACACCATGGCACGGCTGGGAATCGACGTACAAGCCGCCAACA GTCTGTTAAATAACGCCTTCGGTCAGCGGCAAATCTCGACCATTTACCAGCCG ATGAACCAGTATAAAGTGGTGATGGAAGTGGATCCGCGCTATACCCAGGACAT CAGTGCGCTGGAAAAAATGTTCGTTATCAATAACGAAGGCAAAGCGATCCCGC TGTCGTATTTCGCTAAATGGCAACCGGCGAATGCCCCACTATCGGTGAATCAT CAGGGATTATCGGCGGCCTCGACCATTTCGTTTAACCTGCCGACCGGAAAATC GCTCTCGGACGCCAGTGCGGCGATCGATCGCGCAATGACCCAGCTTGGTGTGC CTTCGACGGTGCGCGGCAGTTTTGCCGGCACGGCGCAGGTGTTCCAGGAGACG ATGAACTCGCAGGTGATCCTGATTATCGCCGCCATCGCCACGGTGTATATCGT GCTGGGTATCCTTTACGAGAGTTACGTACATCCGCTGACGATTCTCTCCACCC TGCCCTCGGCGGGCGTTGGAGCGCTGTTGGCGCTGGAGCTGTTCAATGCCCCG TTCAGCCTAATCGCCCTGATAGGGATCATGCTATTAATCGGCATCGTGAAGAA

		AAACGCCATTATGATGGTCGATTTTGCGCTTGAAGCCCAACGGCACGGTAACC TGACGCCGCAGGAAGCTATTTTCCAGGCCTGTCTGCTGCGTTTTCGCCCGATT ATGATGACTACCCTGGCGGCGCTGTTTGGTGCGCTGCCGCTGGTATTGTCGGG CGGCGACGGCTCGGAGCTGCGGCAACCCCTGGGGATCACCATTGTCGGCGGAC TGGTAATGAGCCAGCTCCTTACGCTGTATACCACGCCGGTGGTGTATCTCTTT TTCGACCGTCTGCGGCTGCGTTTTTCGCGTAAACCTAAACAAACGGTAACCGA GTAA
mdtE	1158	>NC_000913.3:3659232-3660389 Escherichia coli str. K-12 substr. MG1655, complete genome ATGAACAGAAGAAGAAAGCTGTTAATACCGTTGTTATTCTGCGGCGCGATGCT CACCGCCTGCGATGACAAATCGGCGGAAAACGCCGCCGCCATGACGCCTGAGG TCGGTGTCGTCACACTCTCCCCCGGTTCGGTCAATGTGTTGAGCGAATTGCCC GGTAGAACCGTTCCTTATGAAGTTGCCGAGATACGTCCCCAGGTGGGCGGTAT TATCATTAAACGCAACTTTATCGAAGGCGATAAAGTGAACCAGGGCGATTCGC TGTATCAGATTGATCCTGCACCTTTACAGGCCGAGCTAAACTCCGCCAAAGGC TCGCTGGCGAAAGCGCTCTCTACCGCCAGCAATGCCCGCATCACCTTTAACCG CCAGGCATCGTTGCTGAAGACCAACTACGTTAGCCGTCAGGATTACGACACCG CGCGCACCCAGTTGAATGAAGCAGAAGCCAATGTCACCGTCGCCAAAGCGGCT GTTGAACAGGCGACGATCAATCTGCAATACGCGAATGTCACCTCGCCGATTAC GGGCGTCAGCGGGAAATCGTCGGTGACCGTCGGCGCACTCGTTACCGCTAATC AGGCAGATTCGCTGGTTACCGTACAACGTCTGGACCCGATTTATGTCGATCTC ACGCAGTCGGTGCAAGATTTCTTACGCATGAAAGAAGAGGTCGCCAGTGGGCA AATCAAACAGGTTCAGGGCAGTACGCCAGTACAGCTCAATCTGGAAAATGGTA AACGCTACAGCCAGACCGGCACGCTGAAATTCTCCGACCCGACAGTGGATGAA ACCACGGGCTCCGTGACGTTACGGGCGATTTTCCCCAACCCAAATGGTGACTT GCTGCCTGGCATGTACGTCACGGCATTAGTGGATGAAGGTAGCCGCCAGAATG TATTACTGGTGCCGCAGGAAGGCGTCACCCACAACGCCCAGGGTAAAGCAACG GCGCTCATTCTGGATAAAGACGATGTCGTGCAGCTACGCGAAATTGAAGCCAG CAAAGCCATCGGCGACCAGTGGGTCGTCACCTCTGGCTTGCAGGCTGGCGATC GGGTGATCGTTTCCGGTTTGCAACGCATTCGTCCGGGTATCAAAGCACGAGCA ATTTCCTCCAGCCAGGAAAACGCCAGCACCGAATCGAAACAATAA
mdtF	3114	>NC_000913.3:3660414-3663527 Escherichia coli str. K-12 substr. MG1655, complete genome ATGGCTAACTATTTTATTGATCGCCCGGTTTTTGCCTGGGTACTTGCCATTAT TATGATGCTTGCAGGTGGTCTGGCGATCATGAACTTACCGGTTGCGCAGTATC CGCAGATTGCGCCACCGACCATTACCGTCAGCGCTACCTATCCAGGTGCCGAT GCGCAAACGGTAGAAGACTCGGTCACTCAGGTGATTGAGCAAAATATGAATGG GCTTGATGGCCTGATGTACATGTCTTCAACCAGTGATGCGGCGGGCAATGCCT CTATCACTCTGACCTTCGAGACTGGGACATCTCCTGATATCGCACAGGTTCAA GTGCAAAATAAACTGCAACTCGCTATGCCTTCATTACCTGAAGCAGTGCAGCA GCAGGGGATTAGCGTCGATAAGTCGAGCAGTAATATCCTGATGGTAGCGGCGT TTATTTCTGATAACGGCAGCCTCAACCAGTACGATATCGCGGACTATGTAGCG TCTAATATCAAAGACCCGCTAAGCCGTACCGCGGGCGTTGGTAGCGTACAACT CTTTGGTTCCGAGTATGCCATGCGTATCTGGCTGGACCCGCAAAAACTCAATA AATATAACCTGGTACCTTCCGATGTTATTTCCCAGATTAAGGTGCAAAACAAC CAGATTTCCGGTGGTCAACTGGGTGGCATGCCACAGGCGGCAGACCAGCAGCT AAACGCCTCGATCATTGTGCAGACGCGTCTGCAAACGCCGGAAGAATTTGGCA AAATCCTGTTGAAAGTTCAGCAAGATGGTTCGCAAGTGCTGCTGCGTGATGTC GCTCGCGTCGAACTTGGGGCGGAAGATTATTCCACCGTGGCACGCTATAACGG CAAACCTGCTGCCGGGATCGCCATCAAACTGGCTGCCGGAGCAAACGCCCTGG ATACCTCGCGGGCAGTCAAAGAGGAACTGAACCGCTTATCAGCCTATTTCCCG GCAAGTCTGAAGACGGTTTATCCTTACGACACCACGCCGTTTATCGAAATTTC TATTCAGGAAGTTTTCAAAACACTGGTTGAGGCTATCATCCTAGTCTTCCTGG TCATGTATCTGTTTTTGCAGAATTTCCGTGCCACAATCATCCCGACGATTGCC GTACCGGTGGTTATTCTCGGGACGTTTGCGATCTTGTCGGCGGTCGGTTTCAC CATCAACACGTTGACTATGTTCGGGATGGTGCTGGCGATAGGGTTACTGGTGG

		ATGACGCCATCGTGGTGGTGGAGAACGTCGAGCGTGTCATTGCGGAAGATAAG CTACCGCCGAAGGAAGCGACGCATAAATCGATGGGGCAGATCCAACGTGCGCT GGTCGGTATTGCCGTTGTTCTTTCCGCAGTGTTTATGCCGATGGCCTTTATGA GCGGTGCAACCGGGGAGATCTACCGCCAGTTCTCCATCACGCTGATCTCCTCC ATGCTGCTTTCAGTATTTGTGGCAATGAGCCTGACCCCTGCCCTGTGCGCCAC CATTCTGAAAGCCGCGCCGGAAGGCGGTCACAAACCTAACGCCCTGTTCGCAC GCTTCAACACGCTGTTTGAAAAATCAACTCAACACTATACCGATAGCACCCGC TCGCTGTTGCGTTGTACCGGTCGCTACATGGTGGTCTACCTGCTGATTTGCGC CGGGATGGCGGTGCTGTTCCTGCGCACGCCGACCTCTTTCTTACCAGAAGAGG ATCAGGGGGTATTTATGACCACCGCGCAGTTACCTTCCGGTGCCACCATGGTT AACACCACGAAAGTGCTGCAACAGGTGACGGATTATTATCTGACTAAAGAGAA AGATAATGTCCAGTCGGTGTTTACCGTTGGCGGCTTTGGCTTCAGCGGTCAGG GGCAAAACAACGGCCTGGCGTTTATCAGTCTCAAGCCGTGGTCTGAACGTGTC GGTGAGGAAAACTCGGTTACCGCGATCATTCAGCGGGCAATGATTGCGTTAAG CAGTATCAATAAAGCCGTCGTCTTCCCGTTCAACTTACCCGCGGTGGCTGAAC TGGGTACCGCGTCAGGTTTTGATATGGAACTGCTGGACAACGGTAACCTGGGG CACGAAAAACTAACCCAGGCGCGAAACGAGCTGTTATCACTGGCAGCGCAATC ACCGAATCAGGTCACCGGGGTACGCCCGAACGGCCTGGAAGATACGCCGATGT TCAAAGTGAACGTCAACGCTGCGAAAGCTGAAGCGATGGGCGTGGCGCTGTCT GATATCAACCAGACAATTTCCACCGCCTTCGGCAGCAGCTACGTGAACGACTT CCTCAACCAGGGGCGGGTGAAAAAAGTGTATGTCCAGGCAGGCACGCCGTTCC GTATGTTGCCGGATAACATCAACCAATGGTATGTACGCAACGCCTCTGGCACG ATGGCACCGCTTTCTGCCTACTCGTCTACCGAATGGACCTATGGTTCACCGCG ACTGGAACGCTACAACGGCATCCCGTCAATGGAGATTTTAGGTGAAGCGGCGG CCGGGAAAAGTACCGGTGACGCCATGAAATTTATGGCAGACCTGGTCGCTAAA CTTCCGGCAGGCGTCGGCTACTCATGGACCGGACTATCGTATCAGGAAGCGTT ATCCTCAAATCAGGCTCCTGCGCTGTATGCGATTTCACTGGTCGTGGTGTTCC TCGCCCTCGCCGCACTCTATGAGAGCTGGTCAATTCCGTTCTCGGTGATGTTG GTTGTTCCGTTAGGCGTCGTTGGCGCATTACTGGCCACCGATCTGCGCGGCTT AAGTAATGACGTCTACTTCCAGGTTGGTTTGCTGACCACCATCGGGCTTTCCG CCAAAAACGCCATCCTGATTGTCGAATTTGCCGTTGAGATGATGCAGAAAGAA GGGAAAACGCCGATAGAGGCAATCATCGAAGCGGCGCGGATGCGTTTACGCCC AATCCTGATGACCTCTCTGGCCTTTATTCTCGGCGTGCTGCCGCTGGTTATCA GTCATGGTGCCGGTTCTGGCGCGCAAAACGCGGTAGGTACCGGCGTGATGGGC GGGATGTTTGCCGCAACAGTGCTGGCAATTTACTTCGTTCCGGTCTTTTTCGT TGTAGTGGAACATCTCTTTGCCCGCTTTAAAAAAGCGTAA
tolC	1482	>NC_000913.3:3178115-3179596 Escherichia coli str. K-12 substr. MG1655, complete genome ATGAAGAAATTGCTCCCCATTCTTATCGGCCTGAGCCTTTCTGGGTTCAGTTC GTTGAGCCAGGCCGAGAACCTGATGCAAGTTTATCAGCAAGCACGCCTTAGTA ACCCGGAATTGCGTAAGTCTGCCGCCGATCGTGATGCTGCCTTTGAAAAAATT AATGAAGCGCGCAGTCCATTACTGCCACAGCTAGGTTTAGGTGCAGATTACAC CTATAGCAACGGCTACCGCGACGCGAACGGCATCAACTCTAACGCGACCAGTG CGTCCTTGCAGTTAACTCAATCCATTTTTGATATGTCGAAATGGCGTGCGTTA ACGCTGCAGGAAAAAGCAGCAGGGATTCAGGACGTCACGTATCAGACCGATCA GCAAACCTTGATCCTCAACACCGCGACCGCTTATTTCAACGTGTTGAATGCTA TTGACGTTCTTTCCTATACACAGGCACAAAAAGAAGCGATCTACCGTCAATTA GATCAAACCACCCAACGTTTTAACGTGGGCCTGGTAGCGATCACCGACGTGCA GAACGCCCGCGCACAGTACGATACCGTGCTGGCGAACGAAGTGACCGCACGTA ATAACCTTGATAACGCGGTAGAGCAGCTGCGCCAGATCACCGGTAACTACTAT CCGGAACTGGCTGCGCTGAATGTCGAAAACTTTAAAACCGACAAACCACAGCC GGTTAACGCGCTGCTGAAAGAAGCCGAAAAACGCAACCTGTCGCTGTTACAGG CACGCTTGAGCCAGGACCTGGCGCGCGAGCAAATTCGCCAGGCGCAGGATGGT CACTTACCGACTCTGGATTTAACGGCTTCTACCGGGATTTCTGACACCTCTTA TAGCGGTTCGAAAACCCGTGGTGCCGCTGGTACCCAGTATGACGATAGCAATA TGGGCCAGAACAAAGTTGGCCTGAGCTTCTCGCTGCCGATTTATCAGGGCGGA

		ATGGTTAACTCGCAGGTGAAACAGGCACAGTACAACTTTGTCGGTGCCAGCGA GCAACTGGAAAGTGCCCATCGTAGCGTCGTGCAGACCGTGCGTTCCTCCTTCA ACAACATTAATGCATCTATCAGTAGCATTAACGCCTACAAACAAGCCGTAGTT TCCGCTCAAAGCTCATTAGACGCGATGGAAGCGGGCTACTCGGTCGGTACGCG TACCATTGTTGATGTGTTGGATGCGACCACCACGTTGTACAACGCCAAGCAAG AGCTGGCGAATGCGCGTTATAACTACCTGATTAATCAGCTGAATATTAAGTCA GCTCTGGGTACGTTGAACGAGCAGGATCTGCTGGCACTGAACAATGCGCTGAG CAAACCGGTTTCCACTAATCCGGAAAACGTTGCACCGCAAACGCCGGAACAGA ATGCTATTGCTGATGGTTATGCGCCTGATAGCCCGGCACCAGTCGTTCAGCAA ACATCCGCACGCACTACCACCAGTAACGGTCATAACCCTTTCCGTAACTGA

8.2 Sequences of *eef* genes used in Chapter 6

Locus tag	Gene	Length	Gene sequence
DR76_3503	*eefR*	567	>CP009072.1:c3743066-3742500 Escherichia coli ATCC 25922, complete genome ATGACCAGCAAGCTGGAGATACGCCACAAACAGCGCCAGGATGAA ATCATTAACGCCGCCCGTCGGTGTTTTCGTCGCTGCGGATTTCAT GCCGCCAGTATGTCGCAAATTGCCAGCGAAGCTCAGCTTAGCGTG GGTCAGATATATCGTTACTTCGCCAACAAAGACGCCATTATTGAA GAGATGGTTCGTCGTATCATCGACTTTCGTATTGCCCAAATGGAC ATTGACGCGCGTACCGATCACTTCCCGGAAGTCCTTGCCTTGCGT AAATCCTTAAATGAGGACGACGACGCGCTGATGCTGGAAGTGGCT GCAGAAGCCACACGCAACCCGCGGGTTATGGCAATGCTTGAAGAA GCTGACGCCAGAATGTTCGCTAACGGTTGTGCACATATGAAAAGA ATGCATCCCCATCTTAGCGACGAGCACATTCGTTGCTGCGTTGAA GTGTTTGCTACCATGATGGAAGGTACGGTTTATCGTCGTTTAACA CCGCAAAAAAGCGATCCACAACATCTGCAGGAAATATATCAGGAT ATTGTTAGTATGTTGATAAATAAATAA
DR76_3502	*eefA*	1122	>CP009072.1:c3742351-3741230 Escherichia coli ATCC 25922, complete genome ATGAAGTATATAGCAACATCTGTAGTGGCAATGCTGCTCTTATCG GGTTGTGATAATACGCAAAGTAACAATTCATCCCCGTCAGAAACA GAAGTCGGCGTTGTTACGCTCAAATCTCAACCGGTTTCGGTAGTC AGTGAATTAACCGGACGTACCAGTGCTGCGCTCAGTGCTGAAGTA CGTCCGCAGGTTGGGGGAATTATCCAGAAACGCTTATTTAAGGAA GGTGATCTGGTCAAGGCTGGACAGCCTCTCTACCAGATTGATGCG GCCAGTTATCAGGCTGCATGGAATGAAGCCCGGGCAGCATTACAA CAAGCACAGGCACTGGTAAAAGCCGATTGCCAGAAAGCGCAGCGT TATACCCGACTGGTGAAAGAGAACGGTGTTTCACAACAGGATGCT GATGATGCTCAGTCTACCTGTGCACAAGATAAAGCCAGTGTAGAG GCGAAAAAAGCCGCACTGGAAACTGCGCGCATTAATCTTGACTGG ACCACGGTAACCGCACCGATTTCGGGGCGTATTGGCATTTCGTCG GTAACCCCTGGCGCACTGGTGACCGCGTCGCAGGATACAGCGTTA ACGACTATTCGTGGTCTGGATACAATGTATGTCGACCTCACTCGC TCCAGTGTCGATTTATTACGTCTGCGTAAACAGTCACTGGCGACC AACAGTGACACCATGAGCGTCTCACTTATTCTGGAAGATGGCACA ACCTACAGCGAAAAAGGGCGTCTGGAACTCACCGAAGTCGCGGTG GATGAGTCTACCGGTTCGGTGACATTACGGGCAATTTTCCCCAAT CCACAACAGCAGTTATTACCGGGAATGTTTGTTCGCGCTCGTGTC GATGAAGGCGTGATGGAAGACGCTATTCTCGCGCCTCAACAGGGC GTTACGCGCGATGCTAAAGGCAATGCAACTGCGCTGGTGGTGAAT AAAGACAATAAAGTAGAGCAGCGAACGCTCGAAACGGGAGAAACG TATGGTGATAAATGGCTGGTGCTGAACGGCCTGCACAACGGCGAC CGACTGATTGTTGAAGGTTCTGCCAAAGTCACTTCAGGGCAGACC GTCAAGGCTGTTGAAGTTCAGGCTAATGGAGGCAACGCCTGA
DR76_3501	*eefB*	3108	>CP009072.1:c3741230-3738123 Escherichia coli ATCC 25922, complete genome ATGTTTTCGCGCTTTTTTGTTCGCCGCCCGGTCTTTGCCTGGGTT ATCGCCATTTTGATTATGTTGGCGGGGATTCTGGCTATTCGCACA TTGCCCGTCGCGCAATATCCTGACGTTGCACCGCCGACCATTAAA ATTTCAGCCACTTATACTGGTGCTTCTGCCGAAACGCTGGAGAAC AGCGTGACTCAGGTTATCGAGCAACAACTCACTGGGCTTGATAAT TTACTCTATTTCAGCTCAACCAGTAGCTCTGATGGTTCGGTCAGT ATTAATGTGACCTTTGAACAAGGTACCGATCCAGATACTGCACAG GTTCAGGTGCAGAATAAAATTCAGCAGGCGGAGTCGCGCCTACCC AGCGAAGTGCAGCAAACGGGTGTTACGGTGGAGAAATCACAAAGC AACTTTTTGCTGATTGCTGCCGTGTATGACACCACTGACAAAGCC

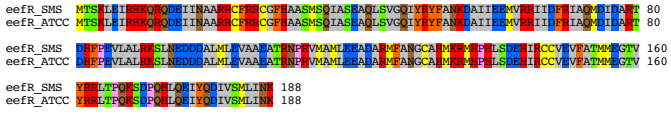
			TCCAGTTCGGATATCGCCGACTGGCTGGTCAGTAACGTTCAGGAC CCCCTGGCGCGTGTTGAAGGTGTGGGGAGTCTGCAAGTCTTTGGC GCGGAATACGCCATGCGCATCTGGCTCGACCCGGCCAAACTGGCG TCTTACTCGCTGATGCCTTCAGACGTGCAAAGTGCTATTGAAGCG CAAAACGTGCAGGTTACTGCCGGGAAAATAGGGGCATTGCCTTCA CCGAATACTCAGCAACTGACCGCAACGGTACGTGCGCAGTCTCGC TTGCAGACGGTGGATCAGTTCAAAAATATTATCGTGAAAAGCCAG TCAGACGGCGCAGTTGTTCGTATAAAAGATGTGGCTCGCGTTGAG ATGGGCAGTGAAGATTATACCGCTATTGGCAAACTTAACGGTCAC CCGTCTGCCGGGGTTGCTGTAATGCTTTCGCCCGGTGCGAATGCG CTGAATACGGCGACGCTGGTCAAGGATAAGATTGCCGAATTCCAG CGGAACATGCCGCAGGGATACGACATCGCATACCCGAAAGACAGC ACTGAATTTATTAAAATCTCCGTAGAGGACGTAATTCAGACGCTG TTTGAAGCCATCGTATTGGTGGTTTGCGTGATGTATTTATTCCTG CAAAACCTGCGTGCCACACTGATTCCGGCGTTAGCTGTCCCCGTC GTTTTGCTGGGCACATTCGGCGTTCTTGCGTTGTTTGGCTATTCG ATTAATACCCTGACGTTGTTTGCAATGGTGCTGGCGATCGGCTTG CTGGTGGACGATGCCATTGTGGTGGTAGAGAACGTCGAACGCATT ATGCGTGACGAAGGGTTGCCCGCGCGTGAAGCCACGGAAAAATCA ATGGGCGAGATTTCTGGCGCACTGGTTGCCATTGCGCTGGTGTTG TCAGCCGTATTCCTGCCGATGGCCTTCTTTGGGGGATCTACGGGC GTAATTTATCGTCAGTTCTCCATCACCATCATCTCCGCAATGCTG CTTTCCGTGGTGGTGGCATTGACCTTGACTCCCGCCCTGTGCGGT TCCGTCCTCCAGCATGTTCCGCCACATAAAAAAGGCTTTTTCGGC GCATTTAACCGCTTTTACCGCCGTACTGAAGATAAATATCAGCGA GGCGTAATTTATGTCCTGCGCCGTGCAGCCCGAACGATGGGGCTT TATGTCGTGCTGGGTGGTGGAATGGCCCTGATGATGTGGAAGCTG CCGGGCAGTTTCTTACCCACTGAAGACCAGGGCGAAATCATGGTG CAGTACACGCTGCCGGCCGGGGCGACCGCTGCCCGTACAGCAGAA GTGAATCGCCAGATTGTTGACTGGTTCCTGATTAACGAGAAAGCA AATACCGATGTCATCTTTACCGTTGATGGTTTCAGTTTTAGCGGC AGCGGACAGAACACCGGGATGGCGTTTGTTTCGTTGAAAAACTGG TCTCAACGTAAAGGGGCAGAAAACACCGCCCAGGCTATCGCCCTA CGGGCTACCAAAGAGCTGGGCACAATTCGTGATGCCACGGTATTT GCGATGACGCCGCCAGCCGTTGATGGGCTGGGGCAAAGCAATGGT TTTACGTTTGAATTGTTAGCTAACGGTGGAACCGATCGTGAAACA CTGCTGCAAATGCGTAATCAGTTGATAGAAAAAGCGAATCAAAGT CCGGAGTTACATTCTGTACGCGCCAATGATTTGCCACAAATGCCG CAATTGCAGGTAGATATTGATAGTAATAAAGCGGTGTCATTAGGG TTGAGTTTGAATGATGTCACCGACACCCTGTCCAGTGCGTGGGGC GGTACTTATGTGAATGACTTTATTGACCGTGGTCGTGTGAAAAAA GTCTACATTCAGGGCGACAGCGAATTTCGTTCCGCGCCGTCAGAC TTAGGTAAATGGTTTGTGCGCGGTAGCGATAACGCCATGACACCA TTCTCTGCTTTTGCGACCACCCGCTGGCTGTATGGACCGGAAAGA CTGGTGCGCTACAACGGCTCGGCAGCCTATGAAATTCAGGGCGAA AACGCGACTGGCTTTAGTTCCGGCGATGCAATGACAAAAATGGAG GAACTGGCAAACAGTCTTCCTGCGGGAACAACCTGGGCCTGGAGT GGTTTGTCATTGCAGGAGAAACTGGCCAGCGGTCAGGCATTAAGT CTGTATGCAGTTTCTATTCTGGTGGTTTTCCTCTGCCTTGCAGCA CTGTACGAAAGCTGGTCAGTCCCGTTCTCGGTCATTCTGGTGATC CCTCTGGGGCTTCTTGGCGCGGCGCTGGCGGCCTGGATGCGTGAT TTAAACAACGACGTTTACTTCCAGGTAGCGCTATTAACTACTATC GGCTTGTCGTCGAAAAACGCCATTCTGATTGTGGAATTTGCTGAA GCGGCGGTTGCAGAGGGCTATTCCTTGAGTCGTGCGGCATTACGC GCGGCGCAGACTCGTTTACGCCCAATCATCATGACCTCGCTGGCG TTTATTGCGGGGGTAATGCCGCTGGCGATAGCAACCGGCGCAGGG GCGAACAGCCGCATCGCCATTGGTACGGGCATTATTGGCGGCACG CTGACCGCTACATTACTGGCCATTTTCTTCGTTCCTCTGTTTTTT

			GTACTGGTGAAACGTTTGTTTGCCGGTAAACCGCGCCGTCAGGAG TAA
DR76_3500	*eefC*	1374	>CP009072.1:c3738119-3736746 Escherichia coli ATCC 25922, complete genome ATGTTGCGTCGTTCTCTTATTTTTCTGGTGTTGTTGAGTGCGGGA TGTGTCTCGCTCGATCCTCACTACAGTACGCCTGAATCCCCAATC CCGGCGACCTTGCCGGGAGCACAGGGCCAGGGGAAGGCAATCAGC CATGATTGGCAGCAGGTGATCCACGATCCCCGATTGCAGCAGGTA GTAACCATTGCACTGAACCGTAACCGCGATGTGCAAAAAGCGATT GCTGATATTGACTCAGCACGGGCGCTCTATGGGCAAACTAATGCG TCGTTATTTCCGACGGTGAATGCGGCCTTAAGTTCGACTCGCAGT CGTTCGCTGGCGAATGGTACGGGAACAACTGCCGAGGCTGACGGC ACGGTGTCCAGTTATACGCTAGATCTGTTTGGTCGTAATCAGAGT TTATCTCGGGCGGCGCGGGAAACCTGGCTTGCCAGTGAATTTACA GCCCAGAACACTCGTTTAACGCTGATTGCTGAAATCAGCACCGCC TGGCTGACACTGGCGGCCGATAACAGCAACCTGGCCCTGGCAAAA GAAACGATGGCCAGTGCAGAAAACTCATTGAAAATTATCCAGCGC CAGCAACAGGTTGGCACGGCGGCAGCGACAGATGTCAGTGAGGCG ATGAGTGTTTATCAGCAAGCGCGCGCCAGCGTTGCCAGCTATCAG ACACAAGTGATGCAGGATAAAAATGCGCTGAACTTACTGGCAGGT ACAACGCTTGCGGAAAATCTGCTACCGGGAACACTGGAAAGCCTG CCGGAACAAATGATAAGTCTGGTTCCTGCGGGAGTATCATCTGAT GTTCTGCTGCGTCGCCCGGATATCCAGGAAGCTGAACATAATTTG AAGAGCGCCAATGCAGATATTGGCGCGGCGCGGGCTAACTTTTTC CCGACGATTTCGCTGACTGCCAGTGCCGGTGTTGGCAGTGACGCA TTGTCGTCGTTGTTCAGCCACGGAATGCAGATCTGGTCATTTGCG CCATCGGTCACATTACCGCTGTTCACTGGCGGAAGTAATCTGGCG CAGCTTCGTTACGCAGAAGCACAAAAGCGTGGGCTCATTGCCACC TATGAAAAAAACGTTCAGAGCGCATTTAAAGATGTCGCTAACGCA CTGGCGCGGCGGACCACGCTTGAAGAGCAACTGGATGCACAGCGC CAGTATGTTAAAGCTGAGCAACAAACGGTCGATGTTGGCTTGCGT CGCTATCAGGCGGGTGTTGGTGATTACCTGACCGTGTTGACGGCG CAGCGCAGCTTGTGGAGCGCACAGCAGGAGCTGCTGGCACTGCAA CTGACTGATTTTACAAACCGAATCACGCTCTGGCAGTCGCTGGGC GGCGGTATGTCATCACTCAAGTAA
DR76_3499	*eefD*	1164	>CP009072.1:c3736737-3735574 Escherichia coli ATCC 25922, complete genome ATGGCTAGAGTCTCTCTTTCATGGGCATTGATTCTTGGTCTTTTA GCCGGTATCGGCCCGATGTGTACCGATCTTTATTTGCCGGCTTTG CCGGAGATGTCTGAGCAACTGGCGGCAACCACGACCATAACGCAA TTAACCCTGACCGCATCACTGATTGGTCTTGGCGTCGGACAACTG TTATTTGGCCCTCTGAGCGACAAAATAGGGCGTAAACGTCCGTTA ATCTTGTCGTTGCTATTGTTTATTGTTTCTTCCATTTTATGCGCG ACAACGAACAATATTTACTGGCTGGTGGTCTGGCGTTTTATTCAA GGGATCGCGGGGGCGGGTGGTTCGGTGCTCTCTCGTTCTATTGCT CGTGACAAATATCAGGGAGTAACGTTGACCCAGTTTTTTGCGCTG TTAATGACGGTGAATGGCCTGGCACCGGTGTTGTCGCCAGTGCTG GGCGGGTACATTGTTAGCACTTTTGACTGGCGCACTTTATTCTGG GTAATGGCTGAAATTAGCACCGTACTGTTGCTGGGCTGCCTGTTA TTTATTAATGAGACCTTGCCAGAAAATAAAAGGGGCTCATCATTG CTATTAACCGGACGAAGCGTGGTGCAGAACCGCCGTTTTATGCGC TTTTGCCTGATTCAAAGCTTTATGCTGGCTGGTTTGTTTGCATAT ATCGGCTCTTCGTCGTTCGTGTTGCAGAAGGAATTTGGCTTTAGT CCAATGCAATTTAGCCTGGTGTTTGGCCTTAACGGCATCGGACTT ATCATTGCTTCATGGATCTTCTCGCGCCTGGCGCGACGGATTAAC GCGATGACATTGTTGCGAGGTGGCCTGATAGCGGCAATTTTGTGT GCATTGCTCACGGTCTTATGCGCATGGATACAATTGCCCATTCCG GCACTGGTGGCATTATTTTTCACCATCGCATTTTGTAGCGGCATC GGCACTGTTGGCGGGGCAGAGGCTATGAGTGCAGTAGGGACGCAG

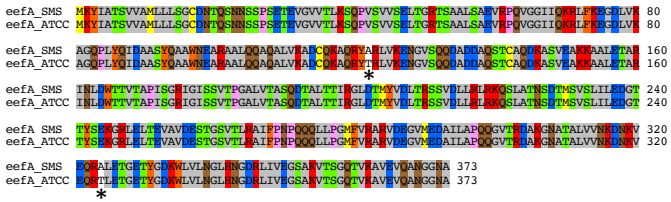
			GAATCTGGAACGGCGTCTGCGTTGATGGGGATGAGCATGTTTGTC TTCGGCGGTATAGCCGCGCCATTGTCGGGAATTGGCGGAGAAACA CTGTTAAAAATGAGTCTGGCAATTACGGTGTGTTATACGCTGGCA TTGCTGGTTGCTCTCACCAGAATCGACAATCAAAAGTAA

8.3 Alignment of *E. coli* SMS-3-5 with *E. coli* ATCC 25922

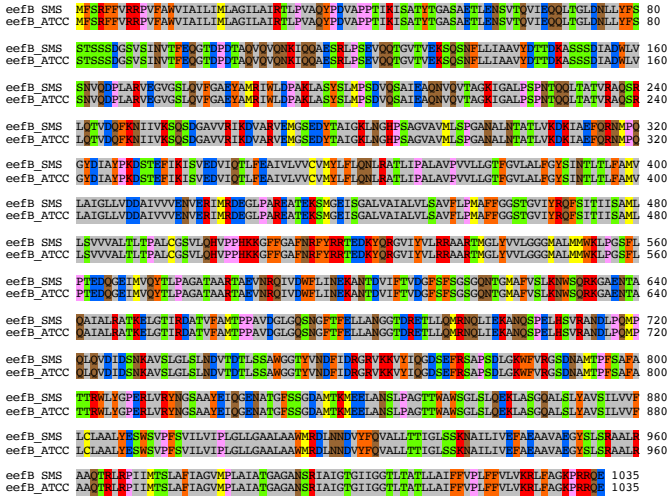
EefR



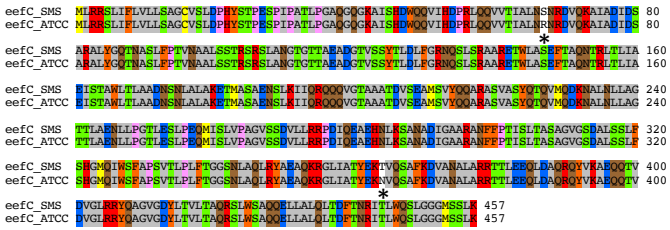
EefA



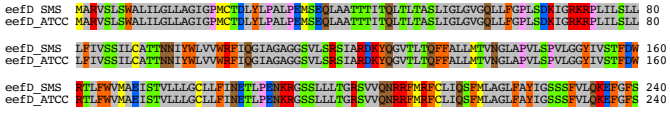
EefB



EefC



EefD



Chapter 9 References

ABRAHAM, E. P. & CHAIN, E. 1940. An Enzyme from Bacteria able to Destroy Penicillin. *Nature,* 146**,** 837.

ABRAM, K., UDAONDO, Z., BLEKER, C., WANCHAI, V., ROBESON II, M. S. & USSERY, D. W. 2020. What can we learn from over 100,000 Escherichia coli genomes?

ABRAM, K., UDAONDO, Z., BLEKER, C., WANCHAI, V., WASSENAAR, T. M., ROBESON, M. S. & USSERY, D. W. 2021. Mash-based analyses of Escherichia coli genomes reveal 14 distinct phylogroups. *Communications Biology,* 4.

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.

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.

ALCOCK, B. P., RAPHENYA, A. R., LAU, T. T. Y., TSANG, K. K., BOUCHARD, E., EDALATMAND, A., HUYNH, W., NGUYEN, A.-L. V., CHENG, A. A., LIU, S., MIN, S. Y., MIROSHNICHENKO, A., TRAN, H.-K., WERFALLI, R. E., NASIR, J. A., OLONI, M., SPEICHER, D. J., FLORESCU, A., SINGH, B., FALTYN, M., HERNANDEZ-KOUTOUCHEVA, A., SHARMA, A. N., BORDELEAU, E., PAWLOWSKI, A. C., ZUBYK, H. L., DOOLEY, D., GRIFFITHS, E., MAGUIRE, F., WINSOR, G. L., BEIKO, R. G., BRINKMAN, F. S. L., HSIAO, W. W. L., DOMSELAAR, G. V. & MCARTHUR, A. G. 2019. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research,* 48.

ALEKSHUN, M. N. & LEVY, S. B. 1999. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol,* 7**,** 410-3.

ALIKHAN, N. F., ZHOU, Z., SERGEANT, M. J. & ACHTMAN, M. 2018. A genomic overview of the population structure of Salmonella. Public Library of Science.

ALON CUDKOWICZ, N. & SCHULDINER, S. 2019. Deletion of the major Escherichia coli multidrug transporter AcrB reveals transporter plasticity and redundancy in bacterial cells. *PLoS One,* 14**,** e0218828.

ANDERSON, G. G., PALERMO, J. J., SCHILLING, J. D., ROTH, R., HEUSER, J. & HULTGREN, S. J. 2003. Intracellular bacterial biofilm-like pods in urinary tract infections. *Science,* 301**,** 105-7.

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.

ARIZA, R. R., COHEN, S. P., BACHHAWAT, N., LEVY, S. B. & DEMPLE, B. 1994. Repressor mutations in the marRAB operon that activate oxidative stress genes and multiple antibiotic resistance in Escherichia coli. *J Bacteriol,* 176**,** 143-8.

BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., DATSENKO, K. A., TOMITA, M., WANNER, B. L. & MORI, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol,* 2**,** 2006 0008.

BARANOVA, N. & NIKAIDO, H. 2002. The baeSR two-component regulatory system activates transcription of the yegMNOB (mdtABCD) transporter gene cluster in Escherichia coli and increases its resistance to novobiocin and deoxycholate. *J Bacteriol,* 184**,** 4168-76.

BARBOSA, T. M. & LEVY, S. B. 2000. The impact of antibiotic use on resistance development and persistence. *Drug Resist Updat,* 3**,** 303-311.

BARNHART, M. M. & CHAPMAN, M. R. 2006. Curli biogenesis and function. *Annu Rev Microbiol,* 60**,** 131-47.

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.

BAY, D. C., ROMMENS, K. L. & TURNER, R. J. 2008. Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochim Biophys Acta,* 1778**,** 1814-38.

BEGHAIN, J., BRIDIER-NAHMIAS, A., NAGARD, H. L., DENAMUR, E. & CLERMONT, O. 2018. ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. *Microbial Genomics,* 4.

BENVENISTE, R. & DAVIES, J. 1973. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc Natl Acad Sci U S A,* 70**,** 2276-80.

BERG, O. G. & KURLAND, C. G. 2002. Evolution of microbial genomes: sequence acquisition and loss. *Mol Biol Evol,* 19**,** 2265-76.

BERGLER, H., FUCHSBICHLER, S., HOGENAUER, G. & TURNOWSKY, F. 1996. The enoyl-[acyl-carrier-protein] reductase (FabI) of Escherichia coli, which catalyzes a key regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA. *Eur J Biochem,* 242**,** 689-94.

BEZABIH, Y. M., SABIITI, W., ALAMNEH, E., BEZABIH, A., PETERSON, G. M., BEZABHE, W. M. & ROUJEINIKOVA, A. 2021. The global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli in the community. *J Antimicrob Chemother,* 76**,** 22-29.

BIELASZEWSKA, M., MELLMANN, A., ZHANG, W., KOCK, R., FRUTH, A., BAUWENS, A., PETERS, G. & KARCH, H. 2011. Characterisation of the Escherichia coli strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis,* 11**,** 671-6.

BINA, X. R., HOWARD, M. F., TAYLOR-MULNEIX, D. L., ANTE, V. M., KUNKLE, D. E. & BINA, J. E. 2018. The Vibrio cholerae RND efflux systems impact virulence factor production and adaptive responses via periplasmic sensor proteins. *PLoS Pathog,* 14**,** e1006804.

BITSORI, M. & GALANAKIS, E. 2019. Treatment of Urinary Tract Infections Caused by ESBL-producing Escherichia coli or Klebsiella pneumoniae. *Pediatr Infect Dis J,* 38**,** e332-e335.

BLAIR, J. M., BAVRO, V. N., RICCI, V., MODI, N., CACCIOTTO, P., KLEINEKATHFER, U., RUGGERONE, P., VARGIU, A. V., BAYLAY, A. J., SMITH, H. E., BRANDON, Y., GALLOWAY, D. & PIDDOCK, L. J. 2015a. AcrB drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity. *Proc Natl Acad Sci U S A,* 112**,** 3511-6.

BLAIR, J. M., SMITH, H. E., RICCI, V., LAWLER, A. J., THOMPSON, L. J. & PIDDOCK, L. J. 2015b. 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.

BLAIR, J. M., WEBBER, M. A., BAYLAY, A. J., OGBOLU, D. O. & PIDDOCK, L. J. 2015c. Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol,* 13**,** 42-51.

BLAIR, J. M. A. & PIDDOCK, L. J. V. 2016. How to Measure Export via Bacterial Multidrug Resistance Efflux Pumps. *mBio,* 7**,** e00840--16.

BLATTNER, F. R., PLUNKETT, G., 3RD, BLOCH, C. A., PERNA, N. T., BURLAND, V., RILEY, M., COLLADO-VIDES, J., GLASNER, J. D., RODE, C. K., MAYHEW, G. F., GREGOR, J., DAVIS, N. W., KIRKPATRICK, H. A., GOEDEN, M. A., ROSE, D. J., MAU, B. & SHAO, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science,* 277**,** 1453-62.

BLIN, K., SHAW, S., STEINKE, K., VILLEBRO, R., ZIEMERT, N., LEE, S. Y., MEDEMA, M. H. & WEBER, T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res,* 47**,** W81-W87.

BOBAY, L.-M. & OCHMAN, H. 2017. The Evolution of Bacterial Genome Architecture. *Frontiers in Genetics,* 8**,** 72-72.

BOHNERT, J. A., SCHUSTER, S., SZYMANIAK-VITS, M. & KERN, W. V. 2011. Determination of real-time efflux phenotypes in Escherichia coli AcrB binding pocket phenylalanine mutants using a 1,2'-dinaphthylamine efflux assay. *PLoS One,* 6**,** e21196.

BOYD, E. F. & HARTL, D. L. 1998. Chromosomal regions specific to pathogenic isolates of Escherichia coli have a phylogenetically clustered distribution. *J Bacteriol,* 180**,** 1159-65.

BRAGA, T. M., MARUJO, P. E., POMBA, C. & LOPES, M. F. S. 2011. Involvement, and dissemination, of the enterococcal small multidrug resistance transporter QacZ in resistance to quaternary ammonium compounds. *Journal of Antimicrobial Chemotherapy,* 66**,** 283-286.

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.

BROCKHURST, M., HARRISON, E., HALL, J., RICHARDS, T., MACNALLY, A. & MACLEAN, C. 2019. The ecology and evolution of pangenomes. *Current Biology*.

BUCHANAN, S. K., SMITH, B. S., VENKATRAMANI, L., XIA, D., ESSER, L., PALNITKAR, M., CHAKRABORTY, R., VAN DER HELM, D. & DEISENHOFER, J. 1999. Crystal structure of the outer membrane active transporter FepA from Escherichia coli. *Nat Struct Biol,* 6**,** 56-63.

BUCKNER, M. M., BLAIR, J. M., LA RAGIONE, R. M., NEWCOMBE, J., DWYER, D. J., IVENS, A. & PIDDOCK, L. J. 2016. Beyond Antimicrobial Resistance: Evidence for a Distinct Role of the AcrD Efflux Pump in Salmonella Biology. *MBio,* 7.

BURSE, A., WEINGART, H. & ULLRICH, M. S. 2004. The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, Erwinia amylovora. *Mol Plant Microbe Interact,* 17**,** 43-54.

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

CAMPRUBI-FONT, C., LOPEZ-SILES, M., FERRER-GUIXERAS, M., NIUBO-CARULLA, L., ABELLA-AMETLLER, C., GARCIA-GIL, L. J. & MARTINEZ-MEDINA, M. 2018. Comparative genomics reveals new single-nucleotide polymorphisms that can assist in identification of adherent-invasive Escherichia coli. *Sci Rep,* 8**,** 2695.

CDC. 2020. *The Final Update on the Multistate Outbreak of E. coli 0157:H7 Infections* [Online]. Altanta, GA. Available: https://www.cdc.gov/media/releases/2020/s0115-ecoli-outbreak.html [Accessed 28/10/2021 2021].

CDC. 2021. *National Shiga toxin-producing Escherichia coli (STEC) Surveillance Annual Report, 2017* [Online]. Atlanta, Georgia: US Department of Health and Human Services, CDC. [Accessed 17/10/2021].

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.

CHANG, A. C. & COHEN, S. N. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol,* 134**,** 1141-56.

CHEN, M., SHI, X., YU, Z., FAN, G., SERYSHEVA, II, BAKER, M. L., LUISI, B. F., LUDTKE, S. J. & WANG, Z. 2021. In situ structure of the AcrAB-TolC efflux pump at subnanometer resolution. *Structure*.

CHEN, T., KOCINSKY, H. S., CHA, B., MURTAZINA, R., YANG, J., TSE, C. M., SINGH, V., COLE, R., ARONSON, P. S., DE JONGE, H., SARKER, R. & DONOWITZ, M. 2015. Cyclic GMP kinase II (cGKII) inhibits NHE3 by altering its trafficking and phosphorylating NHE3 at three required sites: identification of a multifunctional phosphorylation site. *J Biol Chem,* 290**,** 1952-65.

CHENG, Y.-H. H., LIN, T.-L. L., LIN, Y.-T. T. & WANG, J.-T. T. 2018. A putative RND-type efflux pump, H239_3064, contributes to colistin resistance through CrrB in Klebsiella pneumoniae. *Journal of Antimicrobial Chemotherapy,* 73**,** 1509-1516.

CHEREPANOV, P. P. & WACKERNAGEL, W. 1995. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene,* 158**,** 9-14.

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.

CHO, S. H., KIM, J., OH, K. H., HU, J. K., SEO, J., OH, S. S., HUR, M. J., CHOI, Y. H., YOUN, S. K., CHUNG, G. T. & CHOE, Y. J. 2014. Outbreak of enterotoxigenic Escherichia coli O169 enteritis in schoolchildren associated with consumption of kimchi, Republic of Korea, 2012. *Epidemiol Infect,* 142**,** 616-23.

CHOI, U. & LEE, C. R. 2019. Distinct Roles of Outer Membrane Porins in Antibiotic Resistance and Membrane Integrity in Escherichia coli. *Front Microbiol,* 10**,** 953.

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

CLERMONT, O., CHRISTENSON, J. K., DENAMUR, E. & GORDON, D. M. 2013. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep,* 5**,** 58-65.

CLERMONT, O., DIXIT, O. V. A., VANGCHHIA, B., CONDAMINE, B., DION, S., BRIDIER-NAHMIAS, A., DENAMUR, E. & GORDON, D. 2019. Characterization and rapid identification of phylogroup G in Escherichia coli , a lineage with high virulence and antibiotic resistance potential. *Environmental Microbiology,* 21**,** 3107-3117.

CLERMONT, O., GORDON, D. & DENAMUR, E. 2015. Guide to the various phylogenetic classification schemes for Escherichia coli and the correspondence among schemes. *Microbiology,* 161**,** 980-8.
CLSI 2020. *Performance standards for antimicrobial susceptibility testing, 30th edition*, Clinical and Laboratory Standards Institute.
COHEN, S. P., HACHLER, H. & LEVY, S. B. 1993. Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli. *J Bacteriol,* 175**,** 1484-92.
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.
CONANT, G. C. & WOLFE, K. H. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet,* 9**,** 938-50.
CONNOR, C. 2021. *The role of inflammation in the formation of multi-drug resistant lineages of E. coli.* PhD PhD thesis, University of Birmingham.
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*.
COURVALIN, P. 2006. Vancomycin resistance in gram-positive cocci. *Clin Infect Dis,* 42 Suppl 1**,** S25-34.
CRAY, W. C. & MOON, H. W. 1995. Experimental Infection of Calves and Adult Cattle with Escherichia coli O157:H7. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY,* 61**,** 1586-1590.
CROW, A., GREENE, N. P., KAPLAN, E. & KORONAKIS, V. 2017. Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily. 114**,** 12572-12577.
CROXEN, M. A. & FINLAY, B. B. 2010. Molecular mechanisms of Escherichia coli pathogenicity. *Nat Rev Microbiol,* 8**,** 26-38.
CUDKOWICZ, N. A. & SCHULDINER, S. 2019. Deletion of the major Escherichia coli multidrug transporter AcrB reveals transporter plasticity and redundancy in bacterial cells. *PLOS ONE,* 14**,** e0218828-e0218828.
CUI, L., MA, X., SATO, K., OKUMA, K., TENOVER, F. C., MAMIZUKA, E. M., GEMMELL, C. G., KIM, M. N., PLOY, M. C., EL-SOLH, N., FERRAZ, V. & HIRAMATSU, K. 2003. Cell wall thickening is a common feature of vancomycin resistance in Staphylococcus aureus. *J Clin Microbiol,* 41**,** 5-14.
DAGHER, C., SALLOUM, T., ALOUSI, S., ARABAGHIAN, H., ARAJ, G. F. & TOKAJIAN, S. 2018. Molecular characterization of Carbapenem resistant Escherichia coli recovered from a tertiary hospital in Lebanon. *PLoS One,* 13**,** e0203323.
DATSENKO, K. A. & WANNER, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences,* 97**,** 6640-6645.
DE KRAKER, M. E., STEWARDSON, A. J. & HARBARTH, S. 2016. Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Med,* 13**,** e1002184.
DE SOUZA DA-SILVA, A. P., DE SOUSA, V. S., DE ARAÚJO LONGO, L. G., CALDERA, S., BALTAZAR, I. C. L., BONELLI, R. R., SANTORO-LOPES, G., RILEY, L. W. & MOREIRA, B. M. 2020. Prevalence of fluoroquinolone-resistant and broad-spectrum cephalosporin-resistant community-acquired urinary tract infections in Rio de Janeiro: Impact of Escherichia coli genotypes ST69 and ST131. *Infection, Genetics and Evolution,* 85**,** 104452-104452.
DEININGER, K. N., HORIKAWA, A., KITKO, R. D., TATSUMI, R., ROSNER, J. L., WACHI, M. & SLONCZEWSKI, J. L. 2011. A requirement of TolC and MDR efflux pumps for acid adaptation and GadAB induction in Escherichia coli. *PLoS One,* 6**,** e18960.
DELANNOY, S., BEUTIN, L., MARIANI-KURKDJIAN, P., FLEISS, A., BONACORSI, S. & FACH, P. 2017. The Escherichia coli Serogroup O1 and O2 Lipopolysaccharides Are Encoded by Multiple O-antigen Gene Clusters. *Front Cell Infect Microbiol,* 7**,** 30.
DELCOUR, A. H. 2009. Outer membrane permeability and antibiotic resistance. *Biochim Biophys Acta,* 1794**,** 808-16.
DELGADO, M. A., SOLBIATI, J. O., CHIUCHIOLO, M. J., FARIAS, R. N. & SALOMON, R. A. 1999. Escherichia coli outer membrane protein TolC is involved in production of the peptide antibiotic microcin J25. *J Bacteriol,* 181**,** 1968-70.

DENG, Z., SHAN, Y., PAN, Q., GAO, X. & YAN, A. 2013. Anaerobic expression of the gadE-mdtEF multidrug efflux operon is primarily regulated by the two-component system ArcBA through antagonizing the H-NS mediated repression. *Frontiers in Microbiology,* 4**,** 194-194.

DHAMDHERE, G. & ZGURSKAYA, H. I. 2010. Metabolic shutdown in Escherichia coli cells lacking the outer membrane channel TolC. *Mol Microbiol,* 77**,** 743-54.

DU, D., NEUBERGER, A., ORR, M. W., NEWMAN, C. E., HSU, P. C., SAMSUDIN, F., SZEWCZAK-HARRIS, A., RAMOS, L. M., DEBELA, M., KHALID, S., STORZ, G. & LUISI, B. F. 2020. Interactions of a Bacterial RND Transporter with a Transmembrane Small Protein in a Lipid Environment. *Structure,* 28**,** 625-634 e6.

DU, D., WANG, Z., JAMES, N. R., VOSS, J. E., KLIMONT, E., OHENE-AGYEI, T., VENTER, H., CHIU, W. & LUISI, B. F. 2014. Structure of the AcrAB-TolC multidrug efflux pump. *Nature,* 509**,** 512-5.

DU, D., WANG-KAN, X., NEUBERGER, A., VAN VEEN, H. W., POS, K. M., PIDDOCK, L. J. V. V. & LUISI, B. F. 2018. Multidrug efflux pumps: structure, function and regulation. *Nat Rev Microbiol,* 16**,** 523-539.

EDGAR, R. & BIBI, E. 1997. MdfA, an Escherichia coli multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *J Bacteriol,* 179**,** 2274-80.

EDGAR, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-1797.

ELKINS, C. A., MULLIS, L. B., LACHER, D. W. & JUNG, C. M. 2010. Single nucleotide polymorphism analysis of the major tripartite multidrug efflux pump of Escherichia coli: functional conservation in disparate animal reservoirs despite exposure to antimicrobial chemotherapy. *Antimicrob Agents Chemother,* 54**,** 1007-15.

ELKINS, C. A. & NIKAIDO, H. 2002. Substrate specificity of the RND-type multidrug efflux pumps AcrB and AcrD of Escherichia coli is determined predominantly by two large periplasmic loops. *J Bacteriol,* 184**,** 6490-8.

ELLIOTT, S. J., WAINWRIGHT, L. A., MCDANIEL, T. K., JARVIS, K. G., DENG, Y. K., LAI, L. C., MCNAMARA, B. P., DONNENBERG, M. S. & KAPER, J. B. 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic Escherichia coli E2348/69. *Molecular Microbiology,* 28**,** 1-4.

ESCHERICH, T. 1988. The intestinal bacteria of the neonate and breast-fed infant. 1884. *Rev Infect Dis,* 10**,** 1220-5.

ESCOBAR-PÁRAMO, P., CLERMONT, O., BLANC-POTARD, A.-B., BUI, H., BOUGUÉNEC, C. L. & DENAMUR, E. 2004. A Specific Genetic Background Is Required for Acquisition and Expression of Virulence Factors in Escherichia coli. *Molecular Biology and Evolution,* 21**,** 1085-1094.

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.

FISHOVITZ, J., HERMOSO, J. A., CHANG, M. & MOBASHERY, S. 2014. Penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus. *IUBMB Life,* 66**,** 572-7.

FLECKENSTEIN, J. M. & KUHLMANN, F. M. 2019. Enterotoxigenic Escherichia coli Infections. *Curr Infect Dis Rep,* 21**,** 9.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & AL, E. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496-512.

FOONG, W. E., TAM, H. K., CRAMES, J. J., AVERHOFF, B. & POS, K. M. 2019. The chloramphenicol/H+ antiporter CraA of Acinetobacter baumannii AYE reveals a broad substrate specificity. *J Antimicrob Chemother,* 74**,** 1192-1201.

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.

FRANKE, S., GRASS, G., RENSING, C. & NIES, D. H. 2003. Molecular analysis of the copper-transporting efflux system CusCFBA of Escherichia coli. *Journal of Bacteriology,* 185**,** 3804-3812.

FRATAMICO, P. M., DEBROY, C., LIU, Y., NEEDLEMAN, D. S., BARANZONI, G. M. & FENG, P. 2016. Advances in Molecular Serotyping and Subtyping of Escherichia coli. *Front Microbiol,* 7**,** 644.

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*.
FRITZ, S. A. & PURVIS, A. 2010. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits: Selectivity in Extinction Risk. *Conservation Biology,* 24**,** 1042-1051.
FROHLICH, K. S., PAPENFORT, K., BERGER, A. A. & VOGEL, J. 2012. A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res,* 40**,** 3623-40.
GALARDIN, M., KOUMOUTSI, A., HERRERA-DOMINGUEZ, L., VARELA, J. A. C., TELZEROW, A., WAGIH, O., WARTEL, M., CLERMONT, O., DENAMUR, E., TYPAS, A. & BELTRAO, P. 2017. Phenotype prediction in an Escherichia coli strain panel. *bioRxiv*.
GALARDINI, M., KOUMOUTSI, A., HERRERA-DOMINGUEZ, L., VARELA, J. A. C., TELZEROW, A., WAGIH, O., WARTEL, M., CLERMONT, O., DENAMUR, E., TYPAS, A. & BELTRAO, P. 2017. Phenotype inference in an Escherichia coli strain panel. *eLife,* 6.
GARCIA-MENINO, I., GARCIA, V., MORA, A., DIAZ-JIMENEZ, D., FLAMENT-SIMON, S. C., ALONSO, M. P., BLANCO, J. E., BLANCO, M. & BLANCO, J. 2018. Swine Enteric Colibacillosis in Spain: Pathogenic Potential of mcr-1 ST10 and ST131 E. coli Isolates. *Front Microbiol,* 9**,** 2659.
GASCHIGNARD, J., LEVY, C., ROMAIN, O., COHEN, R., BINGEN, E., AUJARD, Y. & BOILEAU, P. 2011. Neonatal Bacterial Meningitis: 444 Cases in 7 Years. *Pediatr Infect Dis J,* 30**,** 212-7.
GAUDU, P., MOON, N. & WEISS, B. 1997. Regulation of the soxRS oxidative stress regulon. Reversible oxidation of the Fe-S centers of SoxR in vivo. *J Biol Chem,* 272**,** 5082-6.
GHOSH, S. & O'CONNOR, T. J. 2017. Beyond Paralogs: The Multiple Layers of Redundancy in Bacterial Pathogenesis. *Front Cell Infect Microbiol,* 7**,** 467.
GIANGROSSI, M., ZATTONI, S., TRAMONTI, A., DE BIASE, D. & FALCONI, M. 2005. Antagonistic role of H-NS and GadX in the regulation of the glutamate decarboxylase-dependent acid resistance system in Escherichia coli. *J Biol Chem,* 280**,** 21498-505.
GILLETTE, W. K., MARTIN, R. G. & ROSNER, J. L. 2000. Probing the Escherichia coli transcriptional activator MarA using alanine-scanning mutagenesis: residues important for DNA binding and activation. *J Mol Biol,* 299**,** 1245-55.
GONZALEZ-ALBA, J. M., BAQUERO, F., CANTÓN, R. & GALÁN, J. C. 2019. Stratified reconstruction of ancestral Escherichia coli diversification. *BMC genomics,* 20**,** 936-936.
GOODALL, E. C. A., ROBINSON, A., JOHNSTON, I. G., JABBARI, S., TURNER, K. A., CUNNINGHAM, A. F., LUND, P. A., COLE, J. A. & HENDERSON, I. R. 2018. The essential genome of Escherichia coli K-12. *mBio,* 9.
GORDON, D. M., CLERMONT, O., TOLLEY, H. & DENAMUR, E. 2008. Assigning Escherichia coli strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environ Microbiol,* 10**,** 2484-96.
GORECKI, K. & MCEVOY, M. M. 2020. Phylogenetic analysis reveals an ancient gene duplication as the origin of the MdtABC efflux pump. *PLoS One,* 15**,** e0228877.
GOULLET, P. & PICARD, B. 1989. Comparative electrophoretic polymorphism of esterases and other enzymes in Escherichia coli. *J Gen Microbiol,* 135**,** 135-43.
GRASS, G. & RENSING, C. 2001. Genes involved in copper homeostasis in Escherichia coli. *J Bacteriol,* 183**,** 2145-7.
GRIFFITH, K. L., FITZPATRICK, M. M., KEEN, E. F., 3RD & WOLF, R. E., JR. 2009. Two functions of the C-terminal domain of Escherichia coli Rob: mediating "sequestration-dispersal" as a novel off-on switch for regulating Rob's activity as a transcription activator and preventing degradation of Rob by Lon protease. *J Mol Biol,* 388**,** 415-30.
HALL, R. J., WHELAN, F. J., MCINERNEY, J. O., OU, Y. & DOMINGO-SANANES, M. R. 2020. Horizontal Gene Transfer as a Source of Conflict and Cooperation in Prokaryotes. *Front Microbiol,* 11**,** 1569.
HANSEN, L. H., JOHANNESEN, E., BURMØLLE, M., SØRENSEN, A. H. & SØRENSEN, S. J. 2004. Plasmid-encoded multidrug efflux pump conferring resistance to olaquindox in Escherichia coli. *Antimicrobial Agents and Chemotherapy,* 48**,** 3332-3337.
HASSAN, K. A., JACKSON, S. M., PENESYAN, A., PATCHING, S. G., TETU, S. G., EIJKELKAMP, B. A., BROWN, M. H., HENDERSON, P. J. & PAULSEN, I. T. 2013. Transcriptomic and biochemical analyses identify a family of chlorhexidine efflux proteins. *Proc Natl Acad Sci U S A,* 110**,** 20254-9.
HASSAN, K. A., LIU, Q., ELBOURNE, L. D. H., AHMAD, I., SHARPLES, D., NAIDU, V., CHAN, C. L., LI, L., HARBORNE, S. P. D., POKHREL, A., POSTIS, V. L. G., GOLDMAN, A., HENDERSON,

P. J. F. & PAULSEN, I. T. 2018. Pacing across the membrane: the novel PACE family of efflux pumps is widespread in Gram-negative pathogens. *Research in Microbiology,* 169**,** 450-454.
HASSAN, K. A., NAIDU, V., EDGERTON, J. R., METTRICK, K. A., LIU, Q., FAHMY, L., LI, L., JACKSON, S. M., AHMAD, I., SHARPLES, D., HENDERSON, P. J. F. & PAULSEN, I. T. 2019. Short-chain diamines are the physiological substrates of PACE family efflux pumps. *Proc Natl Acad Sci U S A,* 116**,** 18015-18020.
HAWKEY, J., ASCHER, D. B., JUDD, L. M., WICK, R. R., KOSTOULIAS, X., CLELAND, H., SPELMAN, D. W., PADIGLIONE, A., PELEG, A. Y. & HOLT, K. E. 2018. Evolution of carbapenem resistance in Acinetobacter baumannii during a prolonged infection. *Microb Genom,* 4.
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.
HAYASHI, T., MAKINO, K., OHNISHI, M., KUROKAWA, K., ISHII, K., YOKOYAMA, K., HAN, C.-G., OHTSUBO, E., NAKAYAMA, K., MURATA, T., TANAKA, M., TOBE, T., IIDA, T., TAKAMI, H., HONDA, T., SASAKAWA, C., OGASAWARA, N., YASUNAGA, T., KUHARA, S., SHIBA, T., HATTORI, M., SHINAGAWA, H. & MAKINO, K. 2001. Complete Genome Sequence of Enterohemorrhagic Eschelichia coli O157:H7 and Genomic Comparison with a Laboratory Strain K-12.
HE, G. X., THORPE, C., WALSH, D., CROW, R., CHEN, H., KUMAR, S. & VARELA, M. F. 2011. EmmdR, a new member of the MATE family of multidrug transporters, extrudes quinolones from Enterobacter cloacae. *Arch Microbiol,* 193**,** 759-65.
HEATH, R. J. & ROCK, C. O. 1995. Enoyl-Acyl Carrier Protein Reductase (fabI) Plays a Determinant Role in Completing Cycles of Fatty Acid Elongation in Escherichia coli. *Journal of Biological Chemistry,* 270**,** 26538-26542.
HEBBELSTRUP JENSEN, B., OLSEN, K. E., STRUVE, C., KROGFELT, K. A. & PETERSEN, A. M. 2014. Epidemiology and clinical manifestations of enteroaggregative Escherichia coli. *Clin Microbiol Rev,* 27**,** 614-30.
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.
HINCHLIFFE, P., YANG, Q. E., PORTAL, E., YOUNG, T., LI, H., TOOKE, C. L., CARVALHO, M. J., PATERSON, N. G., BREM, J., NIUMSUP, P. R., TANSAWAI, U., LEI, L., LI, M., SHEN, Z., WANG, Y., SCHOFIELD, C. J., MULHOLLAND, A. J., SHEN, J., FEY, N., WALSH, T. R. & SPENCER, J. 2017. Insights into the Mechanistic Basis of Plasmid-Mediated Colistin Resistance from Crystal Structures of the Catalytic Domain of MCR-1. *Sci Rep,* 7**,** 39392.
HIRAKAWA, H., INAZUMI, Y., MASAKI, T., HIRATA, T. & YAMAGUCHI, A. 2004. Indole induces the expression of multidrug exporter genes in Escherichia coli. *Molecular Microbiology,* 55**,** 1113-1126.
HOBBS, E. C., YIN, X., PAUL, B. J., ASTARITA, J. L. & STORZ, G. 2012. Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *Proc Natl Acad Sci U S A,* 109**,** 16696-701.
HOBMAN, J. L., PENN, C. W. & PALLEN, M. J. 2007. Laboratory strains of Escherichia coli: Model citizens or deceitful delinquents growing old disgracefully?
HONG, B. K., WANG, M., CHI, H. P., KIM, E. C., JACOBY, G. A. & HOOPER, D. C. 2009. oqxAB encoding a multidrug efflux pump in human clinical isolates of Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy,* 53**,** 3582-3584.
HORESH, G., BLACKWELL, G. A., TONKIN-HILL, G., CORANDER, J., HEINZ, E. & THOMSON, N. R. 2021. A comprehensive and high-quality collection of Escherichia coli genomes and their genes. *Microbial Genomics,* 7**,** 499-499.
HORIYAMA, T. & NISHINO, K. 2014. AcrB, AcrD, and MdtABC Multidrug Efflux Systems Are Involved in Enterobactin Export in Escherichia coli. *PLoS ONE,* 9**,** e108642-e108642.
HOU, P. F., CHEN, X. Y., YAN, G. F., WANG, Y. P. & YING, C. M. 2012. Study of the correlation of imipenem resistance with efflux pumps AdeABC, AdeIJK, AdeDE and AbeM in clinical isolates of Acinetobacter baumannii. *Chemotherapy,* 58**,** 152-8.
HOUSDEN, N. G., WEBBY, M. N., LOWE, E. D., EL-BABA, T. J., KAMINSKA, R., REDFIELD, C., ROBINSON, C. V. & KLEANTHOUS, C. 2021. Toxin import through the antibiotic efflux channel TolC. *Nature Communications 2021 12:1,* 12**,** 1-12.

HYLAND, R. M., SUN, J., GRIENER, T. P., MULVEY, G. L., KLASSEN, J. S., DONNENBERG, M. S. & ARMSTRONG, G. D. 2008. The bundlin pilin protein of enteropathogenic Escherichia coli is an N-acetyllactosamine-specific lectin. *Cellular Microbiology,* 10**,** 177-187.

ICHINOSE, Y., NISHIMURA, T., HARADA, M., KASHIWAGI, R., YAMAMOTO, M., NOUTOSHI, Y., TOYODA, K., TAGUCHI, F., TAKEMOTO, D. & MATSUI, H. 2020. Role of Two Sets of RND-Type Multidrug Efflux Pump Transporter Genes, mexAB-oprM and mexEF-oprN, in Virulence of Pseudomonas syringae pv. tabaci 6605. *Plant Pathol J,* 36**,** 148-156.

JAROCKI, V. M., HESS, S., ANANTANAWAT, K., BERENDONK, T. U. & DJORDJEVIC, S. P. 2021. Multidrug-Resistant Lineage of Enterotoxigenic Escherichia coli ST182 With Serotype O169:H41 in Airline Waste. *Front Microbiol,* 12**,** 731050.

JENSEN, K. F. 1993. The Escherichia coli K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *J Bacteriol,* 175**,** 3401-7.

JEONG, H., KIM, J. S., SONG, S., SHIGEMATSU, H., YOKOYAMA, T., HYUN, J. & HA, N. C. 2016. Pseudoatomic Structure of the Tripartite Multidrug Efflux Pump AcrAB-TolC Reveals the Intermeshing Cogwheel-like Interaction between AcrA and TolC. *Structure,* 24**,** 272-6.

JERSE, A. E., SHARMA, N. D., SIMMS, A. N., CROW, E. T., SNYDER, L. A. & SHAFER, W. M. 2003. A gonococcal efflux pump system enhances bacterial survival in a female mouse model of genital tract infection. *Infect Immun,* 71**,** 5576-82.

JIANG, J. H., HASSAN, K. A., BEGG, S. L., RUPASINGHE, T. W. T., NAIDU, V., PEDERICK, V. G., KHORVASH, M., WHITTALL, J. J., PATON, J. C., PAULSEN, I. T., MCDEVITT, C. A., PELEG, A. Y. & EIJKELKAMP, B. A. 2019. Identification of Novel Acinetobacter baumannii Host Fatty Acid Stress Adaptation Strategies. *mBio,* 10.

KAILASAN VANAJA, S., BERGHOLZ, T. M. & WHITTAM, T. S. 2009. Characterization of the Escherichia coli O157:H7 Sakai GadE regulon. *J Bacteriol,* 191**,** 1868-77.

KAPER, J. B., NATARO, J. P. & MOBLEY, H. L. 2004. Pathogenic Escherichia coli. *Nat Rev Microbiol,* 2**,** 123-40.

KARAOSMANOGLU, K., SAYAR, N. A., KURNAZ, I. A. & AKBULUT, B. S. 2014. Assessment of berberine as a multi-target antimicrobial: a multi-omics study for drug discovery and repositioning. *OMICS,* 18**,** 42-53.

KATZ, L. S., GRISWOLD, T., MORRISON, S. S., CARAVAS, J. A., ZHANG, S., DEN BAKKER, H. C., DENG, X. & CARLETON, H. A. 2019. Mashtree: a rapid comparison of whole genome sequence files *Journal of Open Source Software,* 4.

KAUFFMANN, F. 1947. The serology of the coli group. *J Immunol,* 57**,** 71-100.

KAWAMURA-SATO, K., SHIBAYAMA, K., HORII, T., IIMUMA, Y., ARAKAWA, Y. & OHTA, M. 1999. Role of multiple efflux pumps in Escherichia coli in indole expulsion. *FEMS Microbiol Lett,* 179**,** 345-52.

KIKUCHI, T., MIZUNOE, Y., TAKADE, A., NAITO, S. & YOSHIDA, S.-I. 2005. Curli Fibers Are Required for Development of Biofilm Architecture in Escherichia coli K-12 and Enhance Bacterial Adherence to Human Uroepithelial Cells. *Microbiology and Immunology,* 49**,** 875-884.

KIM, H. S., NAGORE, D. & NIKAIDO, H. 2010. Multidrug efflux pump MdtBC of Escherichia coli is active only as a B2C heterotrimer. *J Bacteriol,* 192**,** 1377-86.

KIM, H. S. & NIKAIDO, H. 2012. Different functions of MdtB and MdtC subunits in the heterotrimeric efflux transporter MdtB(2)C complex of Escherichia coli. *Biochemistry,* 51**,** 4188-97.

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.

KIM, T., BAK, G., LEE, J. & KIM, K. S. 2015. Systematic analysis of the role of bacterial Hfq-interacting sRNAs in the response to antibiotics. *J Antimicrob Chemother,* 70**,** 1659-68.

KITTLESON, J. T., LOFTIN, I. R., HAUSRATH, A. C., ENGELHARDT, K. P., RENSING, C. & MCEVOY, M. M. 2006. Periplasmic metal-resistance protein CusF exhibits high affinity and specificity for both CuI and AgI. *Biochemistry,* 45**,** 11096-102.

KOBAYASHI, A., HIRAKAWA, H., HIRATA, T., NISHINO, K. & YAMAGUCHI, A. 2006. Growth phase-dependent expression of drug exporters in Escherichia coli and its contribution to drug tolerance. *J Bacteriol,* 188**,** 5693-703.

KOBAYASHI, N., NISHINO, K. & YAMAGUCHI, A. 2001. Novel macrolide-specific ABC-type efflux transporter in Escherichia coli. *J Bacteriol,* 183**,** 5639-44.

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.

KU, L. C., BOGGESS, K. A. & COHEN-WOLKOWIEZ, M. 2015. Bacterial meningitis in infants. *Clin Perinatol,* 42**,** 29-45, vii-viii.

KUMARASAMY, K. K., TOLEMAN, M. A., WALSH, T. R., BAGARIA, J., BUTT, F., BALAKRISHNAN, R., CHAUDHARY, U., DOUMITH, M., GISKE, C. G., IRFAN, S., KRISHNAN, P., KUMAR, A. V., MAHARJAN, S., MUSHTAQ, S., NOORIE, T., PATERSON, D. L., PEARSON, A., PERRY, C., PIKE, R., RAO, B., RAY, U., SARMA, J. B., SHARMA, M., SHERIDAN, E., THIRUNARAYAN, M. A., TURTON, J., UPADHYAY, S., WARNER, M., WELFARE, W., LIVERMORE, D. M. & WOODFORD, N. 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *Lancet Infect Dis,* 10**,** 597-602.

KUO, C. H., MORAN, N. A. & OCHMAN, H. 2009. The consequences of genetic drift for bacterial genome complexity. *Genome Res,* 19**,** 1450-4.

KURODA, T. & TSUCHIYA, T. 2009. Multidrug efflux transporters in the MATE family. *Biochim Biophys Acta,* 1794**,** 763-8.

KVIST, M., HANCOCK, V. & KLEMM, P. 2008. Inactivation of efflux pumps abolishes bacterial biofilm formation. *Appl Environ Microbiol,* 74**,** 7376-82.

LADOMERSKY, E. & PETRIS, M. J. 2015. Copper tolerance and virulence in bacteria. *Metallomics,* 7**,** 957-64.

LAMATTINA, J. W., DELROSSI, M., UY, K. G., KEUL, N. D., NIX, D. B., NEELAM, A. R. & LANZILOTTA, W. N. 2017. Anaerobic Heme Degradation: ChuY Is an Anaerobilin Reductase That Exhibits Kinetic Cooperativity. *Biochemistry,* 56**,** 845-855.

LAMERS, R. P., CAVALLARI, J. F. & BURROWS, L. L. 2013. The efflux inhibitor phenylalanine-arginine beta-naphthylamide (PAbetaN) permeabilizes the outer membrane of gram-negative bacteria. *PLoS One,* 8**,** e60666.

LAND, M., HAUSER, L., JUN, S.-R., NOOKAEW, I., LEUZE, M. R., AHN, T.-H., KARPINETS, T., LUND, O., KORA, G., WASSENAAR, T., POUDEL, S. & USSERY, D. W. 2015. Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics,* 15**,** 141-161.

LANGEVIN, A. M., EL MEOUCHE, I. & DUNLOP, M. J. 2020. Mapping the Role of AcrAB-TolC Efflux Pumps in the Evolution of Antibiotic Resistance Reveals Near-MIC Treatments Facilitate Resistance Acquisition. *mSphere,* 5.

LAU, S. Y. & ZGURSKAYA, H. I. 2005. Cell division defects in Escherichia coli deficient in the multidrug efflux transporter AcrEF-TolC. *J Bacteriol,* 187**,** 7815-25.

LE DEROUT, J., BONI, I. V., REGNIER, P. & HAJNSDORF, E. 2010. Hfq affects mRNA levels independently of degradation. *BMC Mol Biol,* 11**,** 17.

LEBLANC, S. K., OATES, C. W. & RAIVIO, T. L. 2011. Characterization of the induction and cellular role of the BaeSR two-component envelope stress response of Escherichia coli. *J Bacteriol,* 193**,** 3367-75.

LEE, I. & SUZUKI, C. K. 2008. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. *Biochim Biophys Acta,* 1784**,** 727-35.

LEE, J. G., HAN, D. S., JO, S. V., LEE, A. R., PARK, C. H., EUN, C. S. & LEE, Y. 2019. Characteristics and pathogenic role of adherent-invasive Escherichia coli in inflammatory bowel disease: Potential impact on clinical outcomes. *PLoS One,* 14**,** e0216165.

LESKI, T. A., BANGURA, U., JIMMY, D. H., ANSUMANA, R., LIZEWSKI, S. E., STENGER, D. A., TAITT, C. R. & VORA, G. J. 2013. Multidrug-resistant tet(X)-containing hospital isolates in Sierra Leone. *Int J Antimicrob Agents,* 42**,** 83-6.

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.

LI, B., YIN, F., ZHAO, X., GUO, Y., WANG, W., WANG, P., ZHU, H., YIN, Y. & WANG, X. 2019a. Colistin Resistance Gene mcr-1 Mediates Cell Permeability and Resistance to Hydrophobic Antibiotics. *Front Microbiol,* 10**,** 3015.

LI, J., ZHANG, H., NING, J., SAJID, A., CHENG, G., YUAN, Z. & HAO, H. 2019b. The nature and epidemiology of OqxAB, a multidrug efflux pump. *Antimicrob Resist Infect Control,* 8**,** 44.
LI, X. Z. & NIKAIDO, H. 2009. Efflux-mediated drug resistance in bacteria: an update. *Drugs,* 69**,** 1555-623.
LI, X. Z., PLESIAT, P. & NIKAIDO, H. 2015. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin Microbiol Rev,* 28**,** 337-418.
LIM, J. Y., YOON, J. & HOVDE, C. J. 2010. A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157. *Journal of Microbiology and Biotechnology,* 20**,** 5-14.
LIN, M. F., LIN, Y. Y., TU, C. C. & LAN, C. Y. 2017. Distribution of different efflux pump genes in clinical isolates of multidrug-resistant Acinetobacter baumannii and their correlation with antimicrobial resistance. *Journal of Microbiology, Immunology and Infection,* 50**,** 224-231.
LING, Z., YIN, W., SHEN, Z., WANG, Y., SHEN, J. & WALSH, T. R. 2020. Epidemiology of mobile colistin resistance genes mcr-1 to mcr-9. *J Antimicrob Chemother,* 75**,** 3087-3095.
LIU, Y. Y., WANG, Y., WALSH, T. R., YI, L. X., ZHANG, R., SPENCER, J., DOI, Y., TIAN, G., DONG, B., HUANG, X., YU, L. F., GU, D., REN, H., CHEN, X., LV, L., HE, D., ZHOU, H., LIANG, Z., LIU, J. H. & SHEN, J. 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases,* 16**,** 161-168.
LOK, C. N., HO, C. M., CHEN, R., TAM, P. K., CHIU, J. F. & CHE, C. M. 2008. Proteomic identification of the Cus system as a major determinant of constitutive Escherichia coli silver resistance of chromosomal origin. *J Proteome Res,* 7**,** 2351-6.
LOMOVSKAYA, O. & BOSTIAN, K. A. 2006. Practical applications and feasibility of efflux pump inhibitors in the clinic--a vision for applied use. *Biochem Pharmacol,* 71**,** 910-8.
LÓPEZ, M., BLASCO, L., GATO, E., PEREZ, A., FERNÁNDEZ-GARCIA, L., MARTÍNEZ-MARTINEZ, L., FERNÁNDEZ-CUENCA, F., RODRÍGUEZ-BAÑO, J., PASCUAL, A., BOU, G. & TOMÁS, M. 2017. Response to Bile Salts in Clinical Strains of Acinetobacter baumannii Lacking the AdeABC Efflux Pump: Virulence Associated with Quorum Sensing. *Frontiers in Cellular and Infection Microbiology*.
LV, L., WAN, M., WANG, C., GAO, X., YANG, Q., PARTRIDGE, S. R., WANG, Y., ZONG, Z., DOI, Y., SHEN, J., JIA, P., SONG, Q., ZHANG, Q., YANG, J., HUANG, X., WANG, M. & LIU, J. H. 2020. Emergence of a Plasmid-Encoded Resistance-Nodulation-Division Efflux Pump Conferring Resistance to Multiple Drugs, Including Tigecycline, in Klebsiella pneumoniae. *mBio,* 11.
LYNCH, M. & CONERY, J. S. 2003. The origins of genome complexity. *Science,* 302**,** 1401-4.
MA, D., ALBERTI, M., LYNCH, C., NIKAIDO, H. & HEARST, J. E. 1996. The local repressor AcrR plays a modulating role in the regulation of acrAB genes of Escherichia coli by global stress signals. *Mol Microbiol,* 19**,** 101-12.
MA, D., COOK, D. N., ALBERTI, M., PON, N. G., NIKAIDO, H. & HEARST, J. E. 1995. Genes acrA and acrB encode a stress-induced efflux system of Escherichia coli. *Molecular Microbiology,* 16**,** 45-55.
MA, K. C., MORTIMER, T. D., HICKS, A. L., WHEELER, N. E., SÁNCHEZ-BUSÓ, L., GOLPARIAN, D., TAIAROA, G., RUBIN, D. H. F., WANG, Y., WILLIAMSON, D. A., UNEMO, M., HARRIS, S. R. & GRAD, Y. H. 2020. Increased antibiotic susceptibility in Neisseria gonorrhoeae through adaptation to the cervical environment. *bioRxiv***,** 2020.01.07.896696-2020.01.07.896696.
MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R. N., POTTER, S. C., FINN, R. D. & LOPEZ, R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res,* 47**,** W636-W641.
MANGES, A. R. & JOHNSON, J. R. 2012. Food-borne origins of Escherichia coli causing extraintestinal infections. *Clin Infect Dis,* 55**,** 712-9.
MARSHALL, R. 2018. *Interactinos within the tripartite drug-efflux pumps of gram-negative bacteria.* PhD PhD thesis, University of Birmingham.
MARTINEZ, J. L., SANCHEZ, M. B., MARTINEZ-SOLANO, L., HERNANDEZ, A., GARMENDIA, L., FAJARDO, A. & ALVAREZ-ORTEGA, C. 2009. Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol Rev,* 33**,** 430-49.
MARTINEZ-MARTINEZ, L., PASCUAL, A. & JACOBY, G. A. 1998. Quinolone resistance from a transferable plasmid. *Lancet,* 351**,** 797-9.
MASI, M., PAGÈS, J. M. & PRADEL, E. 2006. Production of the cryptic EefABC efflux pump in Enterobacter aerogenes chloramphenicol-resistant mutants. *Journal of Antimicrobial Chemotherapy*.

MASI, M., PAGÈS, J. M., VILLARD, C. & PRADEL, E. 2005. The eefABC multidrug efflux pump operon is repressed by H-NS in Enterobacter aerogenes. *Journal of Bacteriology*.

MASI, M., SAINT, N., MOLLE, G. & PAGES, J. M. 2007. The Enterobacter aerogenes outer membrane efflux proteins TolC and EefC have different channel properties. *Biochim Biophys Acta,* 1768**,** 2559-67.

MATAMOROS, S., VAN HATTEM, J. M., ARCILLA, M. S., WILLEMSE, N., MELLES, D. C., PENDERS, J., VINH, T. N., THI HOA, N., BOOTSMA, M. C. J., VAN GENDEREN, P. J., GOORHUIS, A., GROBUSCH, M., MOLHOEK, N., OUDE LASHOF, A. M. L., STOBBERINGH, E. E., VERBRUGH, H. A., DE JONG, M. D. & SCHULTSZ, C. 2017. Global phylogenetic analysis of Escherichia coli and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. *Sci Rep,* 7**,** 15364.

MATHIEU, B., SEBASTIEN, H. & MATHIEU, J. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks.

MATSUMOTO, Y., HAYAMA, K., SAKAKIHARA, S., NISHINO, K., NOJI, H., IINO, R. & YAMAGUCHI, A. 2011. Evaluation of multidrug efflux pump inhibitors by a new method using microfluidic channels. *PLoS One,* 6**,** e18547.

MCALEESE, F., PETERSEN, P., RUZIN, A., DUNMAN, P. M., MURPHY, E., PROJAN, S. J. & BRADFORD, P. A. 2005. A novel MATE family efflux pump contributes to the reduced susceptibility of laboratory-derived Staphylococcus aureus mutants to tigecycline. *Antimicrobial Agents and Chemotherapy,* 49**,** 1865-1871.

MCINERNEY, J. O., MCNALLY, A. & O, M. J. 2017. Why prokaryotes have pangenomes. *Nature Publishing Group*.

MCMURRY, L. M., OETHINGER, M. & LEVY, S. B. 1998. Overexpression of marA, soxS, or acrAB produces resistance to triclosan in laboratory and clinical strains of Escherichia coli. *FEMS Microbiol Lett,* 166**,** 305-9.

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 pathogens,* 15**,** e1008101-e1008101.

MEDEMA, M. H., BLIN, K., CIMERMANCIC, P., DE JAGER, V., ZAKRZEWSKI, P., FISCHBACH, M. A., WEBER, T., TAKANO, E. & BREITLING, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research,* 39**,** W339-46.

MELLMANN, A., HARMSEN, D., CUMMINGS, C. A., ZENTZ, E. B., LEOPOLD, S. R., RICO, A., PRIOR, K., SZCZEPANOWSKI, R., JI, Y., ZHANG, W., MCLAUGHLIN, S. F., HENKHAUS, J. K., LEOPOLD, B., BIELASZEWSKA, M., PRAGER, R., BRZOSKA, P. M., MOORE, R. L., GUENTHER, S., ROTHBERG, J. M. & KARCH, H. 2011. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One,* 6**,** e22751.

MELTON-CELSA, A. R. 2014. Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiology Spectrum,* 2.

MERCHEL PIOVESAN PEREIRA, B. & TAGKOPOULOS, I. 2019. Benzalkonium Chlorides: Uses, Regulatory Status, and Microbial Resistance. *Appl Environ Microbiol,* 85.

MESHCHERYAKOV, V. A., BARKER, C. S., KOSTYUKOVA, A. S. & SAMATEY, F. A. 2013. Function of FlhB, a membrane protein implicated in the bacterial flagellar type III secretion system. *PLoS One,* 8**,** e68384.

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

MULCHANDANI, R., BREHMER, C., BUTT, S., VISHRAM, B., HARRISON, M., MARCHANT, E., FERRIS, S., JORGENSEN, F., SMITH, R., GODBOLE, G., JENKINS, C., DALLMAN, T. J., VERLANDER, N. Q., PHIN, N., TODKILL, D., GHARBIA, S. & HAWKER, J. 2021. Outbreak of Shiga toxin-producing Escherichia coli O157 linked with consumption of a fast-food product containing imported cucumbers, United Kingdom, August 2020. *Int J Infect Dis*.

MURAKAMI, S., NAKASHIMA, R., YAMASHITA, E., MATSUMOTO, T. & YAMAGUCHI, A. 2006. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature,* 443**,** 173-9.

MURRAY, G. G. R., CHARLESWORTH, J., MILLER, E. L., CASEY, M. J., LLOYD, C. T., GOTTSCHALK, M., TUCKER, A. W. D., WELCH, J. J. & WEINERT, L. A. 2021. Genome Reduction Is Associated with Bacterial Pathogenicity across Different Scales of Temporal and Ecological Divergence. *Mol Biol Evol,* 38**,** 1570-1579.

NAGAKUBO, S., NISHINO, K., HIRATA, T. & YAMAGUCHI, A. 2002. The putative response regulator BaeR stimulates multidrug resistance of Escherichia coli via a novel multidrug exporter system, MdtABC. *J Bacteriol,* 184**,** 4161-7.

NATARO, J. P. & KAPER, J. B. 1998. Diarrheagenic Escherichia coli. *Clinical Microbiology Reviews,* 11**,** 142-201.

NATARO, J. P., KAPER, J. B., ROBINS-BROWNE, R., PRADO, V., VIAL, P. & LEVINE, M. M. 1987. Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. *Pediatr Infect Dis J,* 6**,** 829-31.

NATARO, J. P., SCALETSKY, I. C., KAPER, J. B., LEVINE, M. M. & TRABULSI, L. R. 1985. Plasmid-mediated factors conferring diffuse and localized adherence of enteropathogenic Escherichia coli. *Infect Immun,* 48**,** 378-83.

NAYLOR, S. W., LOW, J. C., BESSER, T. E., MAHAJAN, A., GUNN, G. J., PEARCE, M. C., MCKENDRICK, I. J., SMITH, D. G. & GALLY, D. L. 2003. Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic Escherichia coli O157:H7 in the bovine host. *Infect Immun,* 71**,** 1505-12.

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.

NENNINGER, A. A., ROBINSON, L. S. & HULTGREN, S. J. 2008. Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF.

NGUYEN, Y. & SPERANDIO, V. 2012. Enterohemorrhagic E. coli (EHEC) pathogenesis. *Front Cell Infect Microbiol,* 2**,** 90.

NICHOLSON, B. A., WANNEMUEHLER, Y. M., LOGUE, C. M., LI, G. & NOLAN, L. K. 2016. Complete Genome Sequence of the Neonatal Meningitis-Causing Escherichia coli Strain NMEC O18. *Genome Announc,* 4.

NIKAIDO, H. & ZGURSKAYA, H. I. 2001. AcrAB and related multidrug efflux pumps of Escherichia coli. *J Mol Microbiol Biotechnol,* 3**,** 215-8.

NISHIKAWA, Y., HANAOKA, M., OGASAWARA, J., MOYER, N. P. & KIMURA, T. 1995. Heat-stable enterotoxin-producing Escherichia coli O169:H41 in Japan. *Emerg Infect Dis,* 1**,** 61.

NISHIKAWA, Y., HELANDER, A., OGASAWARA, J., MOYER, N. P., HANAOKA, M., HASE, A. & YASUKAWA, A. 1998. Epidemiology and properties of heat-stable enterotoxin-producing Escherichia coli serotype O169:H41. *Epidemiol Infect,* 121**,** 31-42.

NISHINO, K., HAYASHI-NISHINO, M. & YAMAGUCHI, A. 2009. H-NS Modulates Multidrug Resistance of Salmonella enterica Serovar Typhimurium by Repressing Multidrug Efflux Genes acrEF ▿. *Antimicrobial Agents and Chemotherapy,* 53**,** 3541-3543.

NISHINO, K., LATIFI, T. & GROISMAN, E. A. 2006. Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. *Molecular Microbiology,* 59**,** 126-141.

NISHINO, K., SENDA, Y. & YAMAGUCHI, A. 2008a. The AraC-family regulator GadX enhances multidrug resistance in Escherichia coli by activating expression of mdtEF multidrug efflux genes. *J Infect Chemother,* 14**,** 23-9.

NISHINO, K., SENDA, Y. & YAMAGUCHI, A. 2008b. CRP regulator modulates multidrug resistance of Escherichia coli by repressing the mdtEF multidrug efflux genes. *The Journal of antibiotics,* 61**,** 120-7.

NISHINO, K., YAMADA, J., HIRAKAWA, H., HIRATA, T. & YAMAGUCHI, A. 2003. Roles of TolC-dependent multidrug transporters of Escherichia coli in resistance to beta-lactams. *Antimicrob Agents Chemother,* 47**,** 3030-3.

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

NISHINO, K. & YAMAGUCHI, A. 2004. Role of Histone-Like Protein H-NS in Multidrug Resistance of Escherichia coli. *Journal of Bacteriology,* 186**,** 1423-1429.

NISHINO, K., YAMASAKI, S., HAYASHI-NISHINO, M. & YAMAGUCHI, A. 2011. Effect of overexpression of small non-coding DsrA RNA on multidrug efflux in Escherichia coli. *J Antimicrob Chemother,* 66**,** 291-6.

NORDMANN, P., CUZON, G. & NAAS, T. 2009. The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. *Lancet Infect Dis,* 9**,** 228-36.

NORMAN, A., HANSEN, L. H., SHE, Q. & SORENSEN, S. J. 2008. Nucleotide sequence of pOLA52: a conjugative IncX1 plasmid from Escherichia coli which enables biofilm formation and multidrug efflux. *Plasmid,* 60**,** 59-74.

NOVOA, D. & CONROY-BEN, O. 2019. The Anaerobic Efflux Pump MdtEF-TolC Confers Resistance to Cationic Biocides. *bioRxiv***,** 570408-570408.

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. *Journal of Medical Microbiology,* 64**,** 630-635.

NÜESCH-INDERBINEN, M. T., BASCHERA, M., ZURFLUH, K., HÄCHLER, H., NÜESCH, H. & STEPHAN, R. 2017. Clonal Diversity, Virulence Potential and Antimicrobial Resistance of Escherichia coli Causing Community Acquired Urinary Tract Infection in Switzerland. *Frontiers in Microbiology,* 0**,** 2334-2334.

O'NEILL, J. 2014. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. London.

OCHMAN, H. & SELANDER, R. K. 1984. Standard reference strains of Escherichia coli from natural populations. *J Bacteriol,* 157**,** 690-3.

OCHMAN, H., WHITTAM, T. S., CAUGANT, D. A. & SELANDER, R. K. 1983. Enzyme polymorphism and genetic population structure in Escherichia coli and Shigella. *J Gen Microbiol,* 129**,** 2715-26.

OGAWA, W., MINATO, Y., DODAN, H., ONISHI, M., TSUCHIYA, T. & KURODA, T. 2015. Characterization of MATE-type multidrug efflux pumps from Klebsiella pneumoniae MGH78578. *PLoS One,* 10**,** e0121619.

OKONECHNIKOV, K., GOLOSOVA, O., FURSOV, M. & TEAM, U. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics,* 28**,** 1166-7.

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 Biology,* 17**,** 132-132.

OTEO, J., DIESTRA, K., JUAN, C., BAUTISTA, V., NOVAIS, A., PEREZ-VAZQUEZ, M., MOYA, B., MIRO, E., COQUE, T. M., OLIVER, A., CANTON, R., NAVARRO, F., CAMPOS, J. & SPANISH NETWORK IN INFECTIOUS PATHOLOGY, P. 2009. Extended-spectrum beta-lactamase-producing Escherichia coli in Spain belong to a large variety of multilocus sequence typing types, including ST10 complex/A, ST23 complex/A and ST131/B2. *Int J Antimicrob Agents,* 34**,** 173-6.

OUTTEN, F. W., HUFFMAN, D. L., HALE, J. A. & O'HALLORAN, T. V. 2001. The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in Escherichia coli. *J Biol Chem,* 276**,** 30670-7.

PADILLA, E., LLOBET, E., DOMÉNECH-SÁNCHEZ, A., MARTÍNEZ-MARTÍNEZ, L., BENGOECHEA, J. A. & ALBERTÍ, S. 2010. Klebsiella pneumoniae AcrAB efflux pump contributes to antimicrobial resistance and virulence. *Antimicrobial Agents and Chemotherapy,* 54**,** 177-183.

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T. G., FOOKES, M., FALUSH, D., KEANE, J. A. & PARKHILL, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3693.

PAGE, A. J., TAYLOR, B., DELANEY, A. J., SOARES, J., SEEMANN, T., KEANE, J. A. & HARRIS, S. R. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom,* 2**,** e000056.

PARK, Y., CHOI, Q., KWON, G. C. & KOO, S. H. 2020. Molecular epidemiology and mechanisms of tigecycline resistance in carbapenem-resistant Klebsiella pneumoniae isolates. *Journal of Clinical Laboratory Analysis,* 34**,** e23506-e23506.

PARKER, A. & GOTTESMAN, S. 2016. Small RNA Regulation of TolC, the Outer Membrane Component of Bacterial Multidrug Transporters. *J Bacteriol,* 198**,** 1101-13.

PASQUA, M., MICHELACCI, V., DI MARTINO, M. L., TOZZOLI, R., GROSSI, M., COLONNA, B., MORABITO, S. & PROSSEDA, G. 2017. The Intriguing Evolutionary Journey of Enteroinvasive E. coli (EIEC) toward Pathogenicity. *Front Microbiol,* 8**,** 2390.

PÉREZ, A., POZA, M., FERNÁNDEZ, A., DEL CARMEN FERNÁNDEZ, M., MALLO, S., MERINO, M., RUMBO-FEAL, S., CABRAL, M. P. & BOU, G. 2012. Involvement of the AcrAB-TolC efflux pump in the resistance, fitness, and virulence of Enterobacter cloacae. *Antimicrobial Agents and Chemotherapy,* 56**,** 2084-2090.

PERRY, J., WAGLECHNER, N. & WRIGHT, G. 2016. The Prehistory of Antibiotic Resistance. *Cold Spring Harb Perspect Med,* 6.

PETTENGILL, E. A., PETTENGILL, J. B. & BINET, R. 2015. Phylogenetic Analyses of Shigella and Enteroinvasive Escherichia coli for the Identification of Molecular Epidemiological Markers: Whole-Genome Comparative Analysis Does Not Support Distinct Genera Designation. *Front Microbiol,* 6**,** 1573.

PETTY, N. K., BEN ZAKOUR, N. L., STANTON-COOK, M., SKIPPINGTON, E., TOTSIKA, M., FORDE, B. M., PHAN, M. D., GOMES MORIEL, D., PETERS, K. M., DAVIES, M., ROGERS, B. A., DOUGAN, G., RODRIGUEZ-BANO, J., PASCUAL, A., PITOUT, J. D., UPTON, M., PATERSON, D. L., WALSH, T. R., SCHEMBRI, M. A. & BEATSON, S. A. 2014. Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A,* 111**,** 5694-9.

PHE. 2020. *Shiga toxin-producing Escherichia coli (STEC) data: 2018* [Online]. UK: UK Government. [Accessed 17/10/2021].

PHILIPPE, N., MAIGRE, L., SANTINI, S., PINET, E., CLAVERIE, J. M., DAVIN-REGLI, A. V., PAGES, J. M. & MASI, M. 2015. In Vivo Evolution of Bacterial Resistance in Two Cases of Enterobacter aerogenes Infections during Treatment with Imipenem. *PLoS One,* 10**,** e0138828.

PICARD, B., GARCIA, J. S., GOURIOU, S., DURIEZ, P., BRAHIMI, N., BINGEN, E., ELION, J. & DENAMUR, E. 1999. The link between phylogeny and virulence in Escherichia coli extraintestinal infection. *Infect Immun,* 67**,** 546-53.

PITOUT, J. D. & LAUPLAND, K. B. 2008. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an emerging public-health concern. *Lancet Infect Dis,* 8**,** 159-66.

PLETZER, D. & WEINGART, H. 2014. Characterization and regulation of the resistance-nodulation-cell division-type multidrug efflux pumps MdtABC and MdtUVW from the fire blight pathogen Erwinia amylovora. *BMC Microbiol,* 14**,** 185.

PRUIMBOOM-BREES, I. M., MORGAN, T. W., ACKERMANN, M. R., NYSTROM, E. D., SAMUEL, J. E., CORNICK, N. A. & MOON, H. W. 2000. Cattle lack vascular receptors for Escherichia coli O157:H7 Shiga toxins. *Proceedings of the National Academy of Sciences,* 97**,** 10325-10329.

RAFFA, R. G. & RAIVIO, T. L. 2002. A third envelope stress signal transduction pathway in Escherichia coli. *Mol Microbiol,* 45**,** 1599-611.

RAMOS, J. L., MARTINEZ-BUENO, M., MOLINA-HENARES, A. J., TERAN, W., WATANABE, K., ZHANG, X., GALLEGOS, M. T., BRENNAN, R. & TOBES, R. 2005. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev,* 69**,** 326-56.

RAMPIONI, G., PILLAI, C. R., LONGO, F., BONDI, R., BALDELLI, V., MESSINA, M., IMPERI, F., VISCA, P. & LEONI, L. 2017. Effect of efflux pump inhibition on Pseudomonas aeruginosa transcriptome and virulence. *Sci Rep,* 7**,** 11392.

RASKO, D. A., ROSOVITZ, M. J., MYERS, G. S. A., MONGODIN, E. F., FRICKE, W. F., GAJER, P., CRABTREE, J., SEBAIHIA, M., THOMSON, N. R., CHAUDHURI, R., HENDERSON, I. R., SPERANDIO, V. & RAVEL, J. 2008. The pangenome structure of Escherichia coli: Comparative genomic analysis of E. coli commensal and pathogenic isolates. *Journal of Bacteriology,* 190**,** 6881-6893.

REZA, A., SUTTON, J. M. & RAHMAN, K. M. 2019. Effectiveness of Efflux Pump Inhibitors as Biofilm Disruptors and Resistance Breakers in Gram-Negative (ESKAPEE) Bacteria. *Antibiotics (Basel),* 8.

RICCI, V., BLAIR, J. M. & PIDDOCK, L. J. 2014. RamA, which controls expression of the MDR efflux pump AcrAB-TolC, is regulated by the Lon protease. *J Antimicrob Chemother,* 69**,** 643-50.

ROCA, I., MARTI, S., ESPINAL, P., MARTÍNEZ, P., GIBERT, I. & VILA, J. 2009. CraA, a Major Facilitator Superfamily Efflux Pump Associated with Chloramphenicol Resistance in Acinetobacter baumannii. *Antimicrobial Agents and Chemotherapy,* 53**,** 4013-4013.

RÖMLING, U. & GALPERIN, M. Y. 2015. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends in Microbiology,* 23**,** 545-557.

ROSENBERG, E. Y., BERTENTHAL, D., NILLES, M. L., BERTRAND, K. P. & NIKAIDO, H. 2003. Bile salts and fatty acids induce the expression of Escherichia coli AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. *Mol Microbiol,* 48**,** 1609-19.

RUTZ, J. M., LIU, J., LYONS, J. A., GORANSON, J., ARMSTRONG, S. K., MCINTOSH, M. A., FEIX, J. B. & KLEBBA, P. E. 1992. Formation of a gated channel by a ligand-specific transport protein in the bacterial outer membrane. *Science,* 258**,** 471-5.

SAHL, J. W., MORRIS, C. R., EMBERGER, J., FRASER, C. M., OCHIENG, J. B., JUMA, J., FIELDS, B., BREIMAN, R. F., GILMOUR, M., NATARO, J. P. & RASKO, D. A. 2015. Defining the phylogenomics of Shigella species: a pathway to diagnostics. *J Clin Microbiol,* 53**,** 951-60.

SAJIB, M. S. I., TANMOY, A. M., HOODA, Y., RAHMAN, H., ANDREWS, J. R., GARRETT, D. O., ENDTZ, H. P., SAHA, S. K. & SAHA, S. 2021. Tracking the Emergence of Azithromycin Resistance in Multiple Genotypes of Typhoidal Salmonella. *mBio,* 12.

SALIPANTE, S. J., ROACH, D. J., KITZMAN, J. O., SNYDER, M. W., STACKHOUSE, B., BUTLER-WU, S. M., LEE, C., COOKSON, B. T. & SHENDURE, J. 2015. Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. *Genome Res,* 25**,** 119-28.

SANTIAGO, A. G., CHEN, T.-Y., GENOVA, L. A., JUNG, W., THOMPSON, A. M. G., MCEVOY, M. M. & CHEN, P. 2017. Adaptor protein mediates dynamic pump assembly for bacterial metal efflux. *Proceedings of the National Academy of Sciences,* 114**,** 6694-6699.

SAUDER, A. B. & KENDALL, M. M. 2018. After the fact(or): Posttranscriptional gene regulation in enterohemorrhagic Escherichia coli O157:H7. American Society for Microbiology.

SAVAGEAU, M. A. 1983. Escherichia coli habitats, cell types, and molecular mechanisms of gene control. *American Naturalist,* 122**,** 732-744.

SCALETSKY, I. C., SILVA, M. L. & TRABULSI, L. R. 1984. Distinctive patterns of adherence of enteropathogenic Escherichia coli to HeLa cells. *Infect Immun,* 45**,** 534-6.

SCHAFFNER, S. H., LEE, A. V., PHAM, M. T. N., KASSAYE, B. B., LI, H., TALLADA, S., LIS, C., LANG, M., LIU, Y., AHMED, N., GALBRAITH, L. G., MOORE, J. P., BISCHOF, K. M., MENKE, C. C. & SLONCZEWSKI, J. L. 2021. Extreme Acid Modulates Fitness Tradeoffs of Multidrug Efflux Pumps MdtEF-TolC and AcrAB-TolC in Escherichia coli K-12. *Applied and Environmental Microbiology*.

SCHUSTER, S., VAVRA, M., GREIM, L. & KERN, W. V. 2021. Exploring the Contribution of the AcrB Homolog MdtF to Drug Resistance and Dye Efflux in a Multidrug Resistant E. coli Isolate. *Antibiotics,* 10**,** 503-503.

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*.

SELANDER, R. K., DOMINIQUE, A. C. & WHITTAM, T. S. 1987. Genetic structure and variation in natural populations of *Escherichia coli*. *In:* NEIDHARDT, F. C. (ed.) *Escherichia Coli and Salmonella: Cellular and Molecular Biology.* Washington, D.C.: American Society for Microbiology.

SERRA, D. O., RICHTER, A. M. & HENGGE, R. 2013. Cellulose as an Architectural Element in Spatially Structured Escherichia coli Biofilms. *Journal of Bacteriology,* 195**,** 5540-5540.

SERVIN, A. L. 2014. Pathogenesis of human diffusely adhering Escherichia coli expressing Afa/Dr adhesins (Afa/Dr DAEC): current insights and future challenges. *Clin Microbiol Rev,* 27**,** 823-69.

SHAFER, W. M., QU, X., WARING, A. J. & LEHRER, R. I. 1998. Modulation of Neisseria gonorrhoeae susceptibility to vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family. *Proc Natl Acad Sci U S A,* 95**,** 1829-33.

SHARMA, A., GUPTA, V. K. & PATHANIA, R. 2019. Efflux pump inhibitors for bacterial pathogens: From bench to bedside. *Indian J Med Res,* 149**,** 129-145.

SHARMA, P., HAYCOCKS, J. R. J., MIDDLEMISS, A. D., KETTLES, R. A., SELLARS, L. E., RICCI, V., PIDDOCK, L. J. V. & GRAINGER, D. C. 2017. The multiple antibiotic resistance operon of enteric bacteria controls DNA repair and outer membrane integrity. *Nat Commun,* 8**,** 1444.

SJUTS, H., VARGIU, A. V., KWASNY, S. M., NGUYEN, S. T., KIM, H. S., DING, X., ORNIK, A. R., RUGGERONE, P., BOWLIN, T. L., NIKAIDO, H., POS, K. M. & OPPERMAN, T. J. 2016. Molecular basis for inhibition of AcrB multidrug efflux pump by novel and powerful pyranopyridine derivatives. *Proc Natl Acad Sci U S A,* 113**,** 3509-14.

SKARSTAD, K., THONY, B., HWANG, D. S. & KORNBERG, A. 1993. A novel binding protein of the origin of the Escherichia coli chromosome. *J Biol Chem,* 268**,** 5365-70.

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.

SØRENSEN, A. H., HANSEN, L. H., JOHANNESEN, E. & SØRENSEN, S. J. 2003. Conjugative plasmid conferring resistance to olaquindox. *Antimicrobial Agents and Chemotherapy,* 47**,** 798-799.

SRINIVASAN, V. B., RAJAMOHAN, G. & GEBREYES, W. A. 2009. Role of AbeS, a novel efflux pump of the SMR family of transporters, in resistance to antimicrobial agents in Acinetobacter baumannii. *Antimicrob Agents Chemother,* 53**,** 5312-6.

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

SUBASHCHANDRABOSE, S., HAZEN, T. H., BRUMBAUGH, A. R., HIMPSL, S. D., SMITH, S. N., ERNST, R. D., RASKO, D. A. & MOBLEY, H. L. 2014. Host-specific induction of Escherichia coli fitness genes during human urinary tract infection. *Proc Natl Acad Sci U S A,* 111**,** 18327-32.

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.

SUN, J., CHEN, C., CUI, C. Y., ZHANG, Y., LIU, X., CUI, Z. H., MA, X. Y., FENG, Y., FANG, L. X., LIAN, X. L., ZHANG, R. M., TANG, Y. Z., ZHANG, K. X., LIU, H. M., ZHUANG, Z. H., ZHOU, S. D., LV, J. N., DU, H., HUANG, B., YU, F. Y., MATHEMA, B., KREISWIRTH, B. N., LIAO, X. P., CHEN, L. & LIU, Y. H. 2019a. Plasmid-encoded tet(X) genes that confer high-level tigecycline resistance in Escherichia coli. *Nat Microbiol,* 4**,** 1457-1464.

SUN, T., LI, X. D., HONG, J., LIU, C., ZHANG, X. L., ZHENG, J. P., XU, Y. J., OU, Z. Y., ZHENG, J. L. & YU, D. J. 2019b. Inhibitory Effect of Two Traditional Chinese Medicine Monomers, Berberine and Matrine, on the Quorum Sensing System of Antimicrobial-Resistant Escherichia coli. *Front Microbiol,* 10**,** 2584.

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 ‡. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY,* 55**,** 921-924.

TAM, H. K., FOONG, W. E., OSWALD, C., HERRMANN, A., ZENG, H. & POS, K. M. 2021. Allosteric drug transport mechanism of multidrug transporter AcrB. *Nat Commun,* 12**,** 3889.

TANABE, M., SZAKONYI, G., BROWN, K. A., HENDERSON, P. J., NIELD, J. & BYRNE, B. 2009. The multidrug resistance efflux complex, EmrAB from Escherichia coli forms a dimer in vitro. *Biochem Biophys Res Commun,* 380**,** 338-42.

TARR, P. I., GORDON, C. A. & CHANDLER, W. L. 2005. Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *Lancet,* 365**,** 1073-1086.

TEELUCKSINGH, T., THOMPSON, L. K. & COX, G. 2020. The Evolutionary Conservation of Escherichia coli Drug Efflux Pumps Supports Physiological Functions. *Journal of Bacteriology,* 202.

TEGOS, G., STERMITZ, F. R., LOMOVSKAYA, O. & LEWIS, K. 2002. Multidrug pump inhibitors uncover remarkable activity of plant antimicrobials. *Antimicrob Agents Chemother,* 46**,** 3133-41.

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

TERLIZZI, M. E., GRIBAUDO, G. & MAFFEI, M. E. 2017. UroPathogenic Escherichia coli (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. *Frontiers in Microbiology,* 0**,** 1566-1566.

TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M. J., DONATI, C., MEDINI, D., WARD, N. L., ANGIUOLI, S. V., CRABTREE, J., JONES, A. L., DURKIN, A. S., DEBOY, R. T., DAVIDSEN, T. M., MORA, M., SCARSELLI, M., ROS, I. M. Y., PETERSON, J. D., HAUSER, C. R., SUNDARAM, J. P., NELSON, W. C., MADUPU, R., BRINKAC, L. M., DODSON, R. J., ROSOVITZ, M. J., SULLIVAN, S. A., DAUGHERTY, S. C., HAFT, D. H., SELENGUT, J., GWINN, M. L., ZHOU, L., ZAFAR, N., KHOURI, H., RADUNE, D., DIMITROV, G., WATKINS, K., O'CONNOR, K. J. B., SMITH, S., UTTERBACK, T. R., WHITE, O., RUBENS, C. E., GRANDI, G., MADOFF, L. C., KASPER, D. L., TELFORD, J. L., WESSELS, M. R., RAPPUOLI, R. & FRASER, C. M. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences,* 102**,** 13950-13955.

THANASSI, D. G., CHENG, L. W. & NIKAIDO, H. 1997. Active efflux of bile salts by Escherichia coli. *Journal of Bacteriology,* 179**,** 2512-2518.

TOOKE, C. L., HINCHLIFFE, P., BRAGGINTON, E. C., COLENSO, C. K., HIRVONEN, V. H. A., TAKEBAYASHI, Y. & SPENCER, J. 2019. beta-Lactamases and beta-Lactamase Inhibitors in the 21st Century. *J Mol Biol,* 431**,** 3472-3500.

TOUCHON, M., HOEDE, C., TENAILLON, O., BARBE, V., BAERISWYL, S., BIDET, P., BINGEN, E., BONACORSI, S., BOUCHIER, C., BOUVET, O., CALTEAU, A., CHIAPELLO, H., CLERMONT, O., CRUVEILLER, S., DANCHIN, A., DIARD, M., DOSSAT, C., EL KAROUI, M., FRAPY, E., GARRY, L., GHIGO, J. M., GILLES, A. M., JOHNSON, J., LE BOUGUÉNEC, C., LESCAT, M., MANGENOT, S., MARTINEZ-JÉHANNE, V., MATIC, I., NASSIF, X., OZTAS, S., PETIT, M. A., PICHON, C., ROUY, Z., RUF, C. S., SCHNEIDER, D., TOURRET, J., VACHERIE, B., VALLENET, D., MÉDIGUE, C., ROCHA, E. P. C. & DENAMUR, E. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genetics,* 5**,** 1000344-1000344.

TOUCHON, M., PERRIN, A., DE SOUSA, J. A. M., VANGCHHIA, B., BURN, S., O'BRIEN, C. L., DENAMUR, E., GORDON, D. & ROCHA, E. P. C. 2020. Phylogenetic background and habitat drive the genetic diversification of Escherichia coli. *PLoS Genetics,* 16**,** e1008866-e1008866.

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

TREANGEN, T. J., ONDOV, B. D., KOREN, S. & PHILLIPPY, A. M. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol,* 15**,** 524.

TSENG, T. T., GRATWICK, K. S., KOLLMAN, J., PARK, D., NIES, D. H., GOFFEAU, A. & SAIER, M. H., JR. 1999. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol,* 1**,** 107-25.

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.

VANGCHHIA, B., ABRAHAM, S., BELL, J. M., COLLIGNON, P., GIBSON, J. S., INGRAM, P. R., JOHNSON, J. R., KENNEDY, K., TROTT, D. J., TURNIDGE, J. D. & GORDON, D. M. 2016. Phylogenetic diversity, antimicrobial susceptibility and virulence characteristics of phylogroup F Escherichia coli in Australia. *Microbiology (Reading),* 162**,** 1904-1912.

WAMPACH, L., HEINTZ-BUSCHART, A., HOGAN, A., MULLER, E. E. L., NARAYANASAMY, S., LACZNY, C. C., HUGERTH, L. W., BINDL, L., BOTTU, J., ANDERSSON, A. F., DE BEAUFORT, C. & WILMES, P. 2017. Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Front Microbiol,* 8**,** 738.

WANG, D. & FIERKE, C. A. 2013. The BaeSR regulon is involved in defense against zinc toxicity in E. coli. *Metallomics,* 5**,** 372-83.

WANG, Q., ZHANG, P., ZHAO, D., JIANG, Y., ZHAO, F., WANG, Y., LI, X., DU, X. & YU, Y. 2018. Emergence of tigecycline resistance in Escherichia coli co-producing MCR-1 and NDM-5 during tigecycline salvage treatment. *Infect Drug Resist,* 11**,** 2241-2248.

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. 2018. *The role of the AcrB multidrug resistance efflux pump in the biology of Salmonella enterica serovar Typhimurium*

. PhD PhD thesis, University of Birmingham.

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.

WESTON, E. J., PONDO, T., LEWIS, M. M., MARTELL-CLEARY, P., MORIN, C., JEWELL, B., DAILY, P., APOSTOL, M., PETIT, S., FARLEY, M., LYNFIELD, R., REINGOLD, A., HANSEN, N. I., STOLL, B. J., SHANE, A. L., ZELL, E. & SCHRAG, S. J. 2011. The burden of invasive early-onset neonatal sepsis in the United States, 2005-2008. *Pediatr Infect Dis J,* 30**,** 937-41.

WHELAN, F. J., RUSILOWICZ, M., MCINERNEY, J. O., JANE WHELAN, F. & OSCAR MCINERNEY, J. 2020. Coinfinder: detecting significant associations and dissociations in pangenomes. *Microbial Genomics,* 6**,** e000338-e000338.

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.

WIJETUNGE, D. S., GONGATI, S., DEBROY, C., KIM, K. S., COURAUD, P. O., ROMERO, I. A., WEKSLER, B. & KARIYAWASAM, S. 2015. Characterizing the pathotype of neonatal meningitis causing Escherichia coli (NMEC). *BMC Microbiol,* 15**,** 211.

WOOD, D. E., LU, J. & LANGMEAD, B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology,* 20**,** 257-257.

WOODCOCK, D. J., KRUSCHE, P., STRACHAN, N. J. C., FORBES, K. J., COHAN, F. M., MÉRIC, G. & SHEPPARD, S. K. 2017. Genomic plasticity and rapid host switching can promote the evolution of generalism: a case study in the zoonotic pathogen Campylobacter. *Scientific Reports 2017 7:1,* 7**,** 1-13.

WOODFORD, N., TURTON, J. F. & LIVERMORE, D. M. 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev,* 35**,** 736-55.

WRIGHT, G. D. 2005. Bacterial resistance to antibiotics: enzymatic degradation and modification. *Adv Drug Deliv Rev,* 57**,** 1451-70.

WU, J., DUNHAM, W. R. & WEISS, B. 1995. Overproduction and physical characterization of SoxR, a [2Fe-2S] protein that governs an oxidative response regulon in Escherichia coli. *J Biol Chem,* 270**,** 10323-7.

WU, J. & WEISS, B. 1991. Two divergently transcribed genes, soxR and soxS, control a superoxide response regulon of Escherichia coli. *J Bacteriol,* 173**,** 2864-71.

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

XU, Y., LEE, M., MOELLER, A., SONG, S., YOON, B. Y., KIM, H. M., JUN, S. Y., LEE, K. & HA, N. C. 2011. Funnel-like hexameric assembly of the periplasmic adapter protein in the tripartite multidrug efflux pump in gram-negative bacteria. *J Biol Chem,* 286**,** 17910-20.

XUE, Y., DAVIS, A. V., BALAKRISHNAN, G., STASSER, J. P., STAEHLIN, B. M., FOCIA, P., SPIRO, T. G., PENNER-HAHN, J. E. & O'HALLORAN, T. V. 2008. Cu(I) recognition via cation-pi and methionine interactions in CusF. *Nat Chem Biol,* 4**,** 107-9.

YAMADA, J., YAMASAKI, S., HIRAKAWA, H., HAYASHI-NISHINO, M., YAMAGUCHI, A. & NISHINO, K. 2010. Impact of the RNA chaperone Hfq on multidrug resistance in Escherichia coli. *J Antimicrob Chemother,* 65**,** 853-8.

YAMAGUCHI, A., NAKASHIMA, R. & SAKURAI, K. 2015. Structural basis of RND-type multidrug exporters. *Front Microbiol,* 6**,** 327.

YAMAJI, R., RUBIN, J., THYS, E., FRIEDMAN, C. R. & RILEY, L. W. 2018. Persistent Pandemic Lineages of Uropathogenic Escherichia coli in a College Community from 1999 to 2017. *J Clin Microbiol,* 56.

YAMANAKA, H., KOBAYASHI, H., TAKAHASHI, E. & OKAMOTO, K. 2008. MacAB Is Involved in the Secretion of Escherichia coli Heat-Stable Enterotoxin II. *Journal of Bacteriology,* 190**,** 7693-7693.

YAMASAKI, S., NAGASAWA, S., HAYASHI-NISHINO, M., YAMAGUCHI, A. & NISHINO, K. 2011. AcrA dependency of the AcrD efflux pump in Salmonella enterica serovar Typhimurium. *J Antibiot (Tokyo),* 64**,** 433-7.

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, L., SHI, H., ZHANG, L., LIN, X., WEI, Y., JIANG, H. & ZENG, Z. 2021. Emergence of two AcrB substitutions conferring multidrug resistance to Salmonella spp. *Antimicrob Agents Chemother*.

YANG, W., MOORE, I. F., KOTEVA, K. P., BAREICH, D. C., HUGHES, D. W. & WRIGHT, G. D. 2004. TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. *J Biol Chem,* 279**,** 52346-52.

YANG, Z.-K., LUO, H., ZHANG, Y., WANG, B. & GAO, F. 2019. Pan-genomic analysis provides novel insights into the association of E.coli with human host and its minimal genome. *Bioinformatics,* 35**,** 1987-1991.

YASUFUKU, T., SHIGEMURA, K., SHIRAKAWA, T., MATSUMOTO, M., NAKANO, Y., TANAKA, K., ARAKAWA, S., KINOSHITA, S., KAWABATA, M. & FUJISAWA, M. 2011. Correlation of overexpression of efflux pump genes with antibiotic resistance in Escherichia coli Strains clinically isolated from urinary tract infection patients. *J Clin Microbiol,* 49**,** 189-94.

YOON, E. J., COURVALIN, P. & GRILLOT-COURVALIN, C. 2013. RND-type efflux pumps in multidrug-resistant clinical isolates of Acinetobacter baumannii: major role for AdeABC overexpression and AdeRS mutations. *Antimicrob Agents Chemother,* 57**,** 2989-95.

ZGURSKAYA, H. I., WEEKS, J. W., NTREH, A. T., NICKELS, L. M. & WOLLOSCHECK, D. 2015. Mechanism of coupling drug transport reactions located in two different membranes. *Frontiers in Microbiology,* 0**,** 100-100.

ZGURSKAYA, H. I., YAMADA, Y., TIKHONOVA, E. B., GE, Q. & KRISHNAMOORTHY, G. 2009. Structural and functional diversity of bacterial membrane fusion proteins. *Biochim Biophys Acta,* 1794**,** 794-807.

ZHANG, C. Z., CHANG, M. X., YANG, L., LIU, Y. Y., CHEN, P. X. & JIANG, H. X. 2018. Upregulation of AcrEF in Quinolone Resistance Development in Escherichia coli When AcrAB-TolC Function Is Impaired. *Microbial Drug Resistance,* 24**,** 18-23.

ZHANG, Y., XIAO, M., HORIYAMA, T., ZHANG, Y., LI, X., NISHINO, K. & YAN, A. 2011. The multidrug efflux pump MdtEF protects against nitrosative damage during the anaerobic respiration in Escherichia coli. *J Biol Chem,* 286**,** 26576-84.

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

ZOHRA, T., NUMAN, M., IKRAM, A., SALMAN, M., KHAN, T., DIN, M., SALMAN, M., FAROOQ, A., AMIR, A. & ALI, M. 2021. Cracking the Challenge of Antimicrobial Drug Resistance with CRISPR/Cas9, Nanotechnology and Other Strategies in ESKAPE Pathogens. *Microorganisms,* 9.