Understanding the Virulence and Resistance of *Escherichia coli* in Different Countries

Abdulrahman Abdullah Z Almusallam

A thesis submitted in the fulfilment of the requirement of the degree of

Doctor of Philosophy

Cardiff University

School of Medicine, Department of Infection and Immunity

2023

Acknowledgement

Firstly, I would like to thank my supervisor, Mark A. Toleman, for supporting and guiding me and my research and providing me with such a great opportunity to work under his supervision. Dr. Toleman has been a true role model, and I couldn't achieve this work without his guidance.

Jordan Mathias, my good friend, thank you for your friendship, your passion for scientific research, and your positive energy all the time.

Ian, Kirsty, and Janis deserve a special thank for their constant help. The wonderful people, William, and Mei, with whom I shared lab space.

Qassim University for funding my PhD scholarship and for the 6-months extension due to COVID-19 circumstances.

The little ones, Rudy and Taim, for their naughtiness and their understanding of my absence.

Lastly, a special thanks to Norah, my wife, for loving and supporting me constantly and doing the hard work. Without your presence, my world would never be the same.

Dedication

I dedicate this thesis to the memory of my mom, Hussa Almusallm, who always believed in me. May Allah have mercy on your soul and make your residence heaven.

Posters presentations and awards from data generated during this thesis.

European Conference of Clinical Microbiology and Infectious disease (ECCMID) 2020.

ePoster: No. 3061
Comparison of *Escherichia coli* multilocus sequence typing from the UK, Saudi Arabia and Kazakhstan indicates that UK sepsis rates are directly related to carriage of pathogenic *E. coli* strains in the community.
Abdulrahman Almusallam[1], Jordan Mathias[2], Mark Alexander Toleman*[2], Dmitriy Babenko[3]
[1]Cardiff University, Cardiff, United Kingdom, [2]Cardiff University, Cardiff, United Kingdom, [3]Karaganda Medical University, Karaganda, Kazakhstan
https://markterfolg.de/ESCMID/Abstractbook2020.pdf

ePoster: No 3196
Bacteriophage control the prevalence of *Escherichia coli* ST131 in different countries.
Jordan Mathias*[1], Dmitriy Babenko[2], Abdulrahman Almusallam[1], Refath Farzana[3], Mark Alexander Toleman[1]
[1]Cardiff University, Cardiff, United Kingdom, [2]Karaganda Medical University, Karaganda, Kazakhstan, [1]Cardiff University, Cardiff, United Kingdom.
https://markterfolg.de/ESCMID/Abstractbook2020.pdf

ePoster: No 4137
***bla*$_{CTX-M-15}$ and *bla*$_{CTX-M-14}$ genes in UK *Escherichia coli* are found as often on the chromosome as they are on plasmids.** Abdulrahman Almusallam[1], Dmitriy Babenko[2], Jordan Mathias[1], David Wareham[3], Mark Alexander Toleman*[1] [1]Cardiff University, Cardiff, United Kingdom, [2]Karaganda Medical University, Karaganda, Kazakhstan, [3]Blizard Institute, London, United Kingdom.
https://markterfolg.de/ESCMID/Abstractbook2020.pdf

ePoster: No. 5544
Analysis of *Escherichia coli* phylotypes and known sepsis-causing sequence types in UK sewage reveals a direct link between sepsis rates and carriage of pathogenic sequence types in the community. Mark Alexander Toleman*[1], Jordan Mathias[1], Abdulrahman Almusallam[1], Dmitriy Babenko[2] [1]Cardiff University, Cardiff, United Kingdom, [2]Karaganda Medical University, Karaganda, Kazakhstan.
https://markterfolg.de/ESCMID/Abstractbook2020.pdf

This poster was released by ECCMID as a press release (one of only 12 at the conference)
https://www.eurekalert.org/news-releases/709462

The American Society of Microbiology (ASM) Microbe conference 2020

ePoster: No. 13639
Analysis of *E. coli* in Sewage Indicates Huge Differences in Carriage Rates of Pathogenic Strains between Countries. Almusallam[1], J. Mathias[1], D. Babenko[2], M. Toleman[1];
[1]Cardiff Univ., Cardiff, United Kingdom, [2]Karaganda Med. Univ., Karaganda, Kazakhstan
https://www.abstractsonline.com/pp8/#!/9103/presentation/13659

ePoster: No. 13705
***Escherichia coli* Sepsis Rates in the United Kingdom Closely Match Carriage Rates of Pathogenic *E. coli* Strains in the Community.** Almusallam[1], J. Mathias[1], D. Babenko[2], M. A. Toleman[1]; [1]Cardiff Univ., Cardiff, United Kingdom, [2]Karaganda Med. Univ., Karaganda, Kazakhstan.
https://www.abstractsonline.com/pp8/#!/9103/presentation/13705

ePoster: No. 13786
The *Escherichia Coli* ST131 Resistance Epidemic is Manipulated by Bacteriophages.
J. A. T. Mathias[1], A. Almusallam[1], D. Babenko[2], M. Toleman[1];

[1]Cardiff Univ., Cardiff, United Kingdom, [2]Karaganda State Med. Univ., Karagandy, Kazakhstan.
https://www.abstractsonline.com/pp8/#!/9103/presentation/13786

European Conference of Clinical Microbiology and Infectious disease (ECCMID) 2021.

ePoster: No. 04071

Direct linkage between the prevalence of known common ExPEc *Escherichia coli* sequence types in UK sewage to those causing *Escherichia coli* UTI and sepsis. Jordan Mathias (1), Abdulrahman Almusallam (1), Martin Willames (1, 2), Dmitriy Babenko (3), Mark Alexander Toleman* (1) (1) Cardiff University, United Kingdom, (2) UNIFESP, Brazil, (3) Karaganda Medical University, Kazakhstan.

This poster won an award as a top-rated poster at ECCMID (Top 1% of posters presented at ECCMID 2021)
https://2021.eccmid.org/abstracts/top_rated_eposters/

Published abstracts in the European Microbiology journal

Bacteriophage control the prevalence of *Escherichia coli* ST131 in different countries. Mathias J. Almusallam A, Babenko D and **Toleman MA.** EMJ Microbiol Infect Dis. **(2020);**1[1]:18-19.
https://www.emjreviews.com/therapeutic-area/microbiology-infectious diseases/post/abstract?edition=european&page_from=1&page_to=3

Analysis of *Escherichia coli* phylotypes and known sepsis causing sequence types in UK sewage reveals a direct link between sepsis rates and carriage of pathogenic sequence types in the community. Mathias J. Almusallam A, Babenko D and Toleman MA. EMJ Microbiol & Infect Dis. **(2020);**1[1]:27-28.
https://www.emjreviews.com/microbiology-infectious-diseases/abstract/analysis-of-escherichia-coli-phylotypes-and-known-sepsis-causing-sequence-types-in-uk-sewage-reveals-a-direct-link-between-sepsis-rates-and-carriage-of-pathogenic-sequence-types-in-the-community/

Direct Linkage Between the Prevalence of Known Common Extra-intestinal Pathogenic Escherichia coli Sequence Types in UK Sewage to Those Causing E. coli Urinary Tract Infections and Sepsis Mathias J. Almusallam A, Babenko D and Toleman MA. EMJ Microbiol & Infect Dis. **(2021);** 2[1]:33-35.
https://www.emjreviews.com/microbiology-infectious-diseases/abstract/direct-linkage-between-the-prevalence-of-known-common-extra-intestinal-pathogenic-escherichia-coli-sequence-types-in-uk-sewage-to-those-causing-e-coli-urinary-tract-infections-and-sepsis-j100121/

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
μL	Microliter
AMR	Antimicrobial Resistance
bla	β-Lactamase Gene
bp	Base Pair
CFU	Colony Forming Units
CLSI	The Clinical & Laboratory Standards Institute
dCTP	Deoxycytidine Triphosphate
DNA	Deoxyribonucleic Acid
EAEC	Enteroaggregative *E. coli*
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ESBL	Extended Spectrum β-Lactamase
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal pathogenic *E. coli*
g	Gram
HCCA	a-Cyano-4-hydroxycinnamic acid
HGT	Horizontal Gene Transfer
HUS	Haemolytic Uremic Syndrome
ICE	Integrative and Conjugative Element(s)
InPEC	Intestinal Pathogenic *E. coli*
IS	Insertion Site(s)

L	Liter
LB	Lauria Bertani (agar/broth)
LPS	Lipopolysaccharide
MALDI	Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry
Mb	Megabases
MDR	Multi Drug Resistant
MGW	Molecular-Grade Water
MH	Muller-Hinton Agar
mL	Millilitre
MLEE	Multilocus Enzyme Electrophoresis
MLST	Multilocus Sequence Typing
ng	Nano Gram
OD	Optical Density
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame(s)
PBP	Penicillin-Binding Proteins
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
pH	Potential of Hydrogen
rpm	Revolutions per Minute
RT	Room Temperature
SSC	Saline-Sodium Citrate buffer

ST	Sequence Type
TBE	Tris-Borate-EDTA buffer
Tn	Transposon
tRNA	Transfer RNA
UPEC	Uropathogenic *E. coli*
UTIs	Urinary Tract Infections
UV	Ultra-Violet Light
WGS	Whole Genome Sequence
WHO	World Health Organisation

Thesis Summary

Escherichia coli is the most prototypical bacteria known by laymen, while also being the most extensively studied of the prokaryotes (due in part to its being the genetic workhorse of molecular biology). Yet, it is quite remarkable how there is still so much to learn about this organism. Throughout this report, I hope to demonstrate that there are still merits to be found in macro and micro observations of *E. coli*'s genetics, distribution, and behaviour.

Although each chapter within this report exists as an independent entity, the unifying theme is that of extended-spectrum β-lactamase (ESBL) resistance within *E. coli* and its common occurrence with pathogenic strains. The choice of this theme is simple: ESBLs are the most successful and important resistance mechanisms in *E. coli*.

The importance of ESBLs in *E. coli* cannot be understated. *E. coli* is a common bacterial agent of infection, and upon acquisition of an ESBL, infection can become very difficult to treat. This is firstly because the ESBL provides protection against penicillins and most cephalosporins, which are typically the gold standard for a broad-spectrum treatment against infection and secondly that ESBLs are often accompanied by other resistance mechanisms in dominant individual strains of *E. coli,* such as ST131.

E. coli is such a ubiquitous organism and so historic in its study that many papers have been published using various cohorts of isolates. Within this report exists a comprehensive analysis of past *E. coli* literature, and through this scrutiny, I chose contemporary strains of *E. coli* that are routinely carried in the human gut in several different nations.

I specifically utilised a previously archived library of *E. coli* captured across the UK in 2014 as a part of a wider Public Health England study (n= 300) and sampled from

five different geographic regions across the UK, including both human carriage isolates as well as those causing serious disease. It was these isolates that made up the body of work presented in Chapter 3.

The 2014 study sought to understand the epidemiology of resistant *E. coli* across the UK and genotypically compare isolates found in bacteraemia, faeces, sewage, and meat. Interestingly, the study found that resistant *E. coli* associated with meat was unique and separate from those found in bacteraemia and human carriage. All the isolates captured as a part of this 2014 study were grown on antibiotic-selective media, thereby creating a bias in selection. Additionally, one of the key observations of this study was that the proportions of *E. coli* sequence types between blood isolates, faecal isolates, and sewage isolates were very similar. This led to the natural question of ***what types of E. coli are associated with human carriage when antibiotic selection is removed*?** And since this 2014 study had also shown the strength of sewage as a measure of both faecal carriage and subsequent disease: we utilised sewage across the UK to address these questions. Sampling two sites in South Wales, one in Bristol, and three near London, a library of some 600 *E. coli* isolates was generated (approximately 100 isolates per site) without antibiotic bias and representing common *E. coli* types that are carried by the human population at these different locations. This collection serves as a unique snapshot of *E. coli's* pathogenicity, resistance, and prevalence in the UK. This body of work demonstrates how variable levels of *E. coli* bacteraemia rates across the UK are directly related to higher or lower carriage of pathogenic *E. coli* types by local populations and *E. coli* sepsis rates will likely be ineffective until this is acknowledged.

Simultaneous to the UK sewage work, through my own initiative and our international collaborations, we collected sewage from Kazakhstan, Saudi Arabia, and Bangladesh. The *E. coli* libraries created from these samples were again simple in their design. For each site, sewage samples were grown on selective and non-selective media to get an accurate representation of the *E. coli* associated with human carriage in 2019 from various parts of the world. Where possible, whole genome sequencing was utilised to definitively interrogate these organisms and shed light upon *E. coli's* true phylogenetic composition in the different national locations. Within this body of work, I believe I have provided the most comprehensive analysis of chromosomal and plasmid insertions of $bla_{\text{CTX-M-15}}$ to date for *E. coli*. This has demonstrated the huge importance of the chromosomal carriage of this particular resistance mechanism and has also highlighted the chromosomal carriage of other important resistance mechanisms. Equally, I believe my 2019 cohort of some 600 isolates captured across the UK without selection stands as one of the most complete archives of human *E. coli* carriage to date and has shed an important light on the enigma of the rising *E. coli* sepsis rates in the UK. Finally, the collections of strains from different nations have demonstrated that the *E. coli* types circulating in the different nations can be manipulated using antibiotics to select more pathogenic phylotypes. Thus, the use of antibiotics can not only increase antibiotic resistance rates but also increase pathogenic types leading to increased sepsis rates.

Chapter 1 : General Introduction

1.1 Escherichia coli

Escherichia coli is a Gram-negative, rod shaped, non-sporulating, facultative anaerobic bacterium belonging to the Enterobacterales family. It was originally named as bacterium coli commune when first discovered by Theodor Escherich in 1885. He presented his findings to the Society for Morphology and Physiology in Munich (Shulman et al. 2007), and it was not formally renamed until 1958 (Lapage 1979).

Table 1.1 Taxonomic categorization of *Escherichia coli*

Rank	
Domain	Bacteria
Phylum	Proteobacteria
Class	Gammaproteobacteria
Order	Enterobacteriales
Family	Enterobacteriaceae
Genus	*Escherichia*
Species	*Escherichia coli (E. coli)*

E. coli has been the most intensively studied bacterial organism over the last century, with many studies using the *Escherichia coli* K-12 laboratory strain and derivatives. These studies have advanced the disciplines of genetics, molecular biology, physiology, and biochemistry. However, despite this intense study, the huge genetic variation amongst the *E. coli* species has only come to light in recent years with large sequencing programs that have now sequenced more than 200,000 *E. coli* isolates (Lukjancenko et al. 2010). Moreover, the common strains that are carried by human and animal species in different locations of the world are still poorly studied (Hobman et al. 2007). *E. coli* is both a normal inhabitant of the gut flora and the most isolated Gram-negative bacillus in faeces and human infections (Whitman et al.

1998). It colonises the intestines of most warm-blooded animals, in addition to those of many fish (Guzmán et al. 2004) and reptiles (Ramos et al. 2019). It is also known to be able to survive for long periods of time in soils that are classified as temperate (Ishii et al. 2006). *E. coli* may be present in all mammals, although not always in all individuals of each species (Conway and Cohen 2015). For example, some studies have indicated that humans have a prevalence of 90% (Mitsuoka and Hayakawa 1973; Penders et al. 2006); however, wild animals, birds, and reptiles have lower proportions of gut carriage, with 56%, 23%, and 10%, respectively (Gordon and Cowling 2003). Humans typically have 10^7 to 10^9 colony forming units (cfu) per gramme of faeces, and domestic animals typically have a lower amount, 10^4 to 10^6 cfu (Slantez and Bartley 1957; Mitsuoka and Hayakawa 1973). The importance of *E. coli* comes from its ability to cause a variety of serious illnesses, including food poisoning, urinary tract infections, neonatal meningitis, bone infections, intra-abdominal infections, pneumonia, haemolytic uremic syndrome, septic shock, and bacteraemia in humans (Vogt and Dippold 2005). It is transmitted between humans via contact, including skin-to-skin, mucous membrane-to-mucous membrane, and largely by the faecal–oral route or by the consumption of raw undercooked meat products, milk, or raw vegetables, as well as the drinking of water contaminated by sewage. *E. coli* is a highly successful competitor in the human gut, and it is the most common facultative anaerobe found in human intestinal microflora. Interestingly, one study demonstrated that *E. coli* can occupy a highly specific metabolic niche by exploiting its ability to utilise gluconate in the colon (Sweeney et al. 1996).

James Johnson and colleagues (Johnson et al. 2019) initiated a molecular definition of extraintestinal pathogenic *E. coli* (ExPEC) and uropathogenic *E. coli* (UPEC) by identifying bacterial traits among faecal and clinical isolates separately. Two main

general clinical syndromes can result from infection with one of the *E. coli* pathotypes: urinary tract infections (UTIs) and sepsis. Among *E. coli* isolates causing intestinal disease, there are six well-described categories based on their O and H antigens. The O antigen consists of a repeating polysaccharide chain in the lipopolysaccharide (LPS), a component of the outer membrane, while the H antigen is determined by the flagellum. *E. coli* isolates causing intestinal disease to be divided into various pathotypes including: Enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) (Nataro and Kaper 1998).

Shigella spp. and *E. coli* are genetically very similar (Brenner et al. 1972), yet belong to distinct species. In fact, Shigella are essentially decayed versions of *E. coli*. In clinical and public health settings, distinguishing between *E. coli* and *Shigella* and identifying the specific Shigella species is necessary (Chattaway et al. 2017). Infection with any of the four Shigella species can result in dysentery, but only *S. dysenteriae* type 1 can induce haemolytic uremic syndrome (HUS) because of the Shiga toxin it produces. The kidneys, the heart, and the nervous system can all fail and cause death in patients with HUS. Similarlry, *S. dysenteriae* is a common cause of travelers' diarrhoea in the UK. When compared to *S. sonnei* (49%), *S. flexneri* (39%), and *S. boydii* (7%), *S. dysenteriae* is the species that is often isolated least in England (Terry et al. 2018).

Different approaches can be used to define *E. coli* pathotypes, such as the target organ (like the urinary tract and uropathogenic *E. coli*, or UPEC), the infected host (like a bird and avian pathogenic *E. coli*, or APEC), or the connection between an organ and a host (like cerebrospinal fluid in new-borns and newborn meningitis *E.*

coli, or NMEC). The connection with the targeted organs, the presence of specific genes, or the virulence in an animal model (e.g., extra-intestinal pathogenic *E. coli*, or ExPEC); the pathophysiology caused by the strains (for example, diarrhoea and intestine pathogenic *E. coli*, or InPEC); the presence of a certain gene or genes, either on their own or in combination (for example, Shiga-toxin encoding Stx gene and Shiga-toxin producing *E. coli* [STEC]); intimin-encoding *eae* with or without pili-encoding *bfp* gene(s) and typical or atypical enteropathogenic *E. coli* [tEPEC or aEPEC, respectively]); or a specific phenotype when observed outside of the living host, such as adherence to and invasion of epithelial cells or adherent-invasive *E. coli* (AIEC). A comprehensive listing of the pathotypes that are most frequently encountered, along with the primary features of each, is provided in (**Table 1.2**) (Denamur et al. 2021).

Table 1.2 Main characteristics of the more commonly *E. coli* pathotypes (Denamur et al. 2021)

Pathotype Main	Definition basis	Main strain host	Main virulence genes	Strain phylogenetic background	Main sequence types
ExPEC	Non- intestinal infection, specific genes, animal model	Human, domestic mammals, birds	Genes encoding adhesins, toxins, protectins and iron capture systems	B2 D C F	STc131, STc73, STc95, STc12, STc14, STc69 STc88 STc62
UPEC	Isolated from urine	Human, domestic mammals	*papGII, papGIII*	B2 D	STc131, STc73, STc95, STc12, STc14 STc69
NMEC	Isolated from cerebrospinal fluid of neonates	Human	Genes encoding the K1 antigen, pS88 genes	B2 F	STc95 STc59 STc62
Pneumonia associated *E. coli*	Isolated from lung	Human	*hly, sfa*	B2	STc73 STc127 STc141
APEC	Isolated from birds	Poultry	pColV genes	B2 C G	STc95 STc88 STc117
InPEC	Diarrhoeal disease	Human, domestic mammals	Various	All phylogroups	Numerous
STEC and/or EHEC	*stx* genes	Human, cattle[c], sheep[c]	*stx, eae, ehxA*	E B1	STc11 STc29 ST17
EPEC	Attaching and effacing lesions on intestinal epithelial cells	Human, domestic mammals	*eae, bfp*	A B1 B2 E	ST1788 (EPEC5) STc10 (EPEC10) STc3 (EPEC2) STc328 (EPEC7) STc15 (EPEC1) STc28 (EPEC4) STc5342 (EPEC8) STc2346 (EPEC9) STc335, STc32

Pathotype Main	**Definition basis**	**Main strain host**	**Main virulence genes**	**Strain phylogenetic background**	**Main sequence types**
ETEC	Heat-stable and heat- labile enterotoxins	Human, pig, cattle	Genes encoding enterotoxins and colonization factors	A, B1, C, E	Numerous
EIEC	Colonocyte invasion	Strictly human	*ipa, isc, vir* Inactivation of *nadA, nadB* and *cadA*	A B1 E	ST6 ST270 ST280
EAEC	Aggregative adhesion on enterocytes	Human, domestic mammals	Aggregative adherence fimbriae (*aaf/agg*) and transcriptional (*aggR*) genes	A, B1, B2, D	Numerous
DAEC	Diffuse adhesion on enterocytes	Human	Genes encoding adhesins (*afa* and *dra*)	All phylogroups	Numerous
AIEC	Adhesion and invasion of intestinal epithelial cells	Human	Unknown	All phylogroups with a majority of B2	ST135 ST73, ST95, ST127, ST131
Hybrid InPEC	EHEC and EAEC characteristics	Human	*stx, aggABCD, aggR*	B1	ST678
Hybrid InPEC–ExPEC	HUS and septicaemia	Human, cattle	*stx*, *eae*, *pS88* ExPEC genes	A	ST301

In recent years, complex hybrid pathotypes have emerged, either within the InPEC pathotypes (for example, enterohaemorrhagic *E. coli* (EHEC) and enteroaggregative *E. coli* (EAEC)) or between the InPEC and ExPEC pathotypes. Some examples of these complex hybrid pathotypes include enterohaemorrhagic *E. coli* (EHEC) and enteroaggregative *E. coli* (for example, EHEC and ExPEC), which cause the pathotype categorisation to be complicated and difficult to understand. Additional uncertainty was caused by both the description of cryptic Escherichia clades12 and the fact that it is difficult to identify other Escherichia species. Simultaneously, the accessibility of whole-genome sequencing (WGS) has resulted in the collection of genomic data, which may enable phylogenomic methodologies to be used in the classification of pathogenic *E. coli* strains.

1.2 The population structure of *E. coli*

1.2.1 Escherichia genus and *E. coli* species phylogeny.

The taxonomy of the genus *Escherichia* has undergone significant revisions because of the identification of five previously unknown *Escherichia* clades as well as the reassignment of the species *Escherichia blattae*, *Escherichia hermanii*, and *Escherichia vulneris* to other genera. *Escherichia coli*, *Escherichia albertii*, and *Escherichia fergusonii* are the three species that make up the genus *Escherichia* today. Additionally, there are five *Escherichia* clades that have been labelled I–V. These clades are indistinguishable from *E. coli* in terms of their phenotypic characteristics, yet they differ from *E. coli* in their nucleotide sequences in varying degrees based on the typical identity of the nucleotides. *E. coli* sensu stricto possesses a robust phylogenetic structure that is comprised of at least eight distinct

phylogenetic groups that are organised into two primary clusters: phylogroups B2, G, F, and D; and phylogroups A, B1, C, and E (Denamur et al. 2021).

1.2.2 Order and disorder in *E. coli* genomes

Plasmids and a circular chromosome are the two components that make up the *E. coli* genome. The *E. coli* genome size (excluding *Shigella*) can range anywhere from 4.2 to 6.0 Mb, which is equivalent to 3,900–5,800 genes, respectively (Hendrickson 2009). During the process of divergence, *E. coli* underwent several frequent acquisitions and deletions of DNA segments, which led to the current diversity. The observed size variation is significantly influenced by phylogenetic factors. The shortest genomes are found in strains that belong to phylogroups A and B1, whereas the biggest genomes are found in strains that belong to phylogroup E (Chattaway et al. 2017). The core genome of *E. coli* has around 2,000-2,500 genes, while the remaining genes in a strain come from the pan-genome. All *E. coli* strains share these core genes. The pangenome, also known as the total number of genes among a species, can range anywhere from 15,000 genes in 20 genomes to 90,000 genes in 1,500 genomes, depending on the number of genomes that are analysed (Touchon et al. 2020). On the chromosome, there is evidence of a significant amount of homologous recombination, also known as gene conversion, which occurs at least as frequently as mutations do (Land et al. 2015).

This wide range of reported recombined fragment lengths could be explained by the various effects of the restriction systems of the recipient cells, which reduces the length of the acquired DNA, as well as the result of successive overlapping incorporations of large fragments, which over time leads to a mosaic of small segments. Both factors can shorten the length of the acquired DNA. There are two

hotspots of recombination situated at the O-antigen biosynthetic gene cluster and the *fim* operon. Although recombination is dispersed over the chromosome, it is less common towards the terminal of replication. It is interesting to note that the same regions that are hotspots for recombination are also hotspots for integration. This suggests that homologous recombination may play a role in the acquisition of substantial portions of DNA. In addition, many *E. coli* strains have plasmids, often between two and four different plasmids per strain. The type of plasmid is species-specific as demonstrated by the frequent occurrence of plasmids belonging to the incompatibility group *IncF* and *Inc*I in *E. coli*, and the number of plasmids appears to fluctuate within species according to the strain's phylogeny (Denamur et al. 2021).

1.2.3 Defining a clone in the genomic era

A clone is formed of isolates that are either indistinguishable from one another or extremely closely related and that share a common ancestor. *E. coli* has a clonal population structure, which means that alleles are strongly linked in a way that is not random (linkage disequilibrium) and that only a few of the many possible multilocus genotypes are often found (Smith et al. 1993). As mentioned before, this is because of the rate and pattern of recombination (horizontal transfer), which involves small pieces that do not break up the vertical evolution caused by mutation. By using phenotypic serotype determination (somatic (O-serogroup), capsular (K), and flagellar (H) antigens) (Orskov et al. 1976), and multilocus enzyme electrophoresis (MLEE) (Selander and Levin 1980), this clonal structure was initially identified. The development and widespread use of Sanger sequencing technology has led to the replacement of these phenotypic tests by the sequencing of 500bp segments of seven to eight housekeeping genes. This method is known as multilocus sequence typing (MLST), which is an analogue to MLEE (Maiden et al. 1998). Whole genome

sequencing (WGS) has made it possible to do in-silico typing of strains for a variety of genotyping techniques. It is possible to do traditional typing using it, such as O:H typing, MLST typing (Jaureguy et al. 2008), or *fimH* allele typing (Roer et al. 2017), which is based on small differences in sequence. Core genome MLST or whole genome MLST can also be done, depending on how closely related the isolates are to one another. In addition to allele typing, sequencing yields nucleotide sequences, and a phylogenetic tree may be generated based on SNPs in the core genome that reliably infer the evolutionary history of the isolates. Both processes are performed by sequencing the genome. Currently, MLST is the most common method to define strains. The sequence type designation of an isolate is a clone (Denamur et al. 2021).

1.3 Virulence and differente linages of *E. coli*

E. coli ST131 strains are the leading cause of bacteraemia and urinary tract infections (cystitis and pyelonephritis) acquired in hospitals and communities worldwide. One study used genome sequencing to analyse the temporal and spatial correlation of many *E. coli* ST131 strains obtained from six different geographical areas worldwide. *E. coli* ST131 strains are unique among other extraintestinal pathogenic *E. coli* and originated from a single progenitor strain before the year 2000 (Petty et al. 2014). ST131 is also responsible for other forms of infections, such as intra-abdominal and soft tissue infections, meningitis, osteoarticular infection, myositis, epididymo-orchitis, and septic shock. The range of infections that ST131 causes is typical of ExPEC and has heightened the degree of worry over this strain, which is already seen as a serious potential issue related to its resistance to many drugs (Nicolas-Chanoine et al. 2014). The origin of ST131 is not well defined in the current literature (Loh et al. 2013). However, we have unpublished evidence

indicating that *E. coli* ST131 was widespread across several major hospitals in India in 2000 (Babenko et al. 2019).

In a study that was conducted at a general hospital in San Francisco between the years 2007 and 2010, ST131 *E. coli* infections were found in 23% of *E. coli* bacteraemia patients. However, more than 70% of these infections were found in patients who were admitted within the first 48 hours of the study, which indicates that these infections were acquired in the community (Imai et al. 2019). Within this therapeutically relevant species, *E. coli* sequence type 131 (ST131) has emerged as the most common extraintestinal pathogenic lineage. This lineages connection with fluoroquinolone and extended-spectrum cephalosporin resistance has a substantial influence on therapy. The majority of ESBL-associated *E. coli* infections are typically caused by a recently emerged ExPEC clone of ST131 (ST131RX) that has spread throughout the world in the last two decades (Banerjee and Johnson 2014). The proliferation of ST131 is largely responsible for the rise in antibiotic resistance in ExPEC infections, particularly ESBL-producing and fluoroquinolone-resistant ExPEC (Stoesser et al. 2016). Resistance is mostly mediated by the ESBL genes *bla*$_{CTX\text{-}M\text{-}15}$, *bla*$_{CTX\text{-}M\text{-}27}$ and *bla*$_{CTX\text{-}M\text{-}14}$. Although *bla*$_{CTX\text{-}M\text{-}15}$ is by far the most common of these and typically expressed plasmids. ST131 is also often fluoroquinolone resistant which is mostly caused by mutations in the chromosomal genes *gyrA* and *parC* (Stoesser et al. 2016). Although upregulated efflux pumps and plasmid-encoded resistance mechanisms can reduce fluoroquinolone (FQ) susceptibility in *E. coli*, high-level resistance typically requires one-two point mutations within the quinolone resistance-determining regions (QRDRs) of *gyrA* and *parC*, which are the chromosomal genes encoding the FQ targets DNA gyrase and topoisomerase IV, respectively (Johnson et al. 2013).

Other studies have investigated the genetic basis of non-ST131 ExPEC success. Alhashash et al. studied 22 isolates of multidrug-resistant ST73 and observed significant genomic heterogeneity, indicating an extended period of evolution. The authors determined that the rise in ST73 resistance was not caused by the dissemination of certain clones. ST73 is highly successful, unlike the clonal ST131, however the reasons for its success remain unknown (Alhashash et al. 2016). ST1193 is a newly identified worldwide MDR high-risk strain that significantly contributes to community-acquired urine and bloodstream infections. ST1193 is mimicking *E. coli* ST131, the most successful multidrug-resistant clone in history. Both clones have been crucial in the worldwide dissemination of multidrug-resistant *E. coli*. The global incidence of ST1193 has been on the rise since 2012, and it has even replaced ST131 in some locations. The *bla*$_{\text{CTX-M}}$ genes are quickly spreading among ST1193 isolates, similar to what happened with ST131 in the 2000s (Pitout et al. 2022).

Although main sequence types play a crucial role in worldwide infections, there is still limited understanding of the variations in performance across distinct lineages in extraintestinal sites and the gut.

1.4 Serotyping and subtyping of *E. coli*

1.4.1 Serotyping

The practise of serotyping, which has been around since the 1940s and has now been refined into standard operating procedures, is one of the most frequent methods for categorising *E. coli*. The O: K: H system of serotyping is derived from a mixture of three immunogenic structures. These components are the lipopolysaccharide (LPS) (O antigen), the capsular antigen (K antigen), and the

flagellar (H) antigen. The current serotyping system for *E. coli* includes a total of 188 O groups with the designations O1 through O188. In addition, the groups O31, O47, O67, O72, and O122 have been removed from the scheme (Kauffmann 1947). The scheme consists of a total of 53 H antigens, which are classified from H1 to H56. However, the H13, H22, and H50 antigens have been removed (Joensen et al. 2015). Agglutination responses of the O antigen with antisera produced from rabbits against each of the O groups serve as the basis for this theory (ørskov et al. 1975). It is important to note that serotyping may also be done in silico and based on the sequencing of the entire genome. This is possible since serotyping is dependent on gene products (Orskov et al. 1976).

1.4.2 Phylotyping

There are a few different approaches that may be used to identify the phylotypes of *E. coli* (See materials and methods section). In the 1980s, the research conducted by Whittam and his colleagues demonstrated that the bacterium *E. coli* had a genetic substructure. These studies demonstrate that the distribution of *E. coli* was not random, and that various strains belong to distinct phylogroups depending on the location from where they were isolated (Whittam et al. 1983).

1.4.3 Multilocus enzyme electrophoresis (MLEE)

From the beginning of the 20th century, microbiologists have relied on MLEE as a tried-and-tested technique for a better comprehension of the heterogeneity of microorganisms, particularly virulent bacteria. The first time that MLEE was used to differentiate between different types of bacteria was in 1963 (Andrews and Chilton 1999). The MLEE method distinguishes between strains by analysing the electrophoretic mobility of primary metabolic enzymes. The amino acid sequence of

a protein is what determines its electrostatic charge. Differences in molecular weight and the total charge of the enzymes might show these discrepancies, which in turn cause variances in the mobility of the protein when it is exposed to electrophoresis. Electromorphs are the name given to these mobility variations, and they may be directly correlated with allelic variation at the locus of the gene that they belong to. When a collection of enzymes is analysed at the same time, the varying mobilities of the enzymes form a pattern of protein bands that is specific to each strain and can indicate the multilocus genotype of the strain (Maiden 2006). Malate dehydrogenase, glucose-6-phosphate dehydrogenase, glutamate dehydrogenase, and phosphoglucomutase are some examples of housekeeping enzymes that are investigated in MLEE. These enzymes are all involved in the fundamental metabolic processes of the cell. In 1984, Ochman and Salander utilised MLEE to describe the genome of 72 *E. coli* isolates creating 72 reference strains of *E. coli* that were isolated from a range of hosts and geographical areas. This set was used as a representative of the range of genotypic variation in the species as a whole, and it is known as the *Escherichia coli* reference (ECOR) collection (Selander et al. 1986).

1.4.4 Multilocus sequence typing (MLST)

MLST is a genetic typing method that was initially proposed in 1998. It is a method that is both powerful and portable, and it was designed to characterise bacterial isolates on a molecular level. Because it is based on the population genetic concepts that underpinned the earlier technique of MLEE, this typing method has an advantage over many other approaches in that it easily compares isolates from different nations and is completely transferrable between research groups (Maiden 2006). MLST databases make it possible to compare directly between different bacterial isolates at a reasonable cost. It typically uses seven allele fragments that

are between 400 and 600 base pairs in length (this was because this length was easily sequenced by Sanger sequencing). All genes that were targeted are fundamental components of the genome that are shared by all *E. coli* strains (IE core genome genes). First, PCR is used to amplify each gene, and then the allele number is checked against the list of alleles that has been saved in the different databases. Once additional alleles are identified, they are added to the system and each one is given a distinct sequential number assigned by earliest discovery. As a result, the ST is a combination of the numbers of all the alleles present in each strain. There are currently three MLST scheme for *E. coli*: MLST website hosted at Michigan State University (the USA), Warwick Medical school-now hosted by Enterobase.com (the UK), and Pasteur Institute (France) (**Table 1.3**) (Sahl et al. 2012).

Table 1.3 The main characteristics of three *E. coli* MLST databases

Location	Genes	Website	No. STs	No. strains
Michigan State University	aspC, clpX, fadD, icd, lysP, mdh, uidA	http://www.shigatox.net/ecmlst/cgi-bin/index	1081	3965
Warwick Medical School	adk, fumC, gyrB, icd, mdh, purA, recA	https://enterobase.warwick.ac.uk/	>10,000	>10,000
Pasteur Institute	dinB, icd, pabB, polB, putP, trpA, trpB, uidA	http://www.pasteur.fr/recherche/genopole/PF8/mlst/E. coli.html	771	1311

The Warwick *E. coli* MLST scheme is the most widely used scheme and is used in our studies. The MLST method itself does not give any information on phylotype. Therefore, we use both MLST and phylogroup methods to fully describe the *E. coli* isolates collected through this study.

1.5 ExPEC *E. coli*

Urinary tract infections (also known as UTIs) are among the most prevalent bacterial infections found in the community both in developing and developed nations. They afflict around 150 million individuals on a yearly basis all over the world.

UTIs are a major cause of morbidity in baby boys, older men, and females of all ages (Stamm and Norrby 2001). UTIs can be caused by a variety of microorganisms, including bacteria (both Gram-negative and Gram-positive) and fungi. Uropathogenic *Escherichia coli* (ExPEC, previously called UPEC) are the most prevalent cause of bacterial urinary tract infections (Flores-Mireles et al. 2015). It can lead to several serious complications, including recurring infections, pyelonephritis with sepsis, premature delivery, and damage to the kidneys in children. In addition, the expansion of antibiotic resistance in *E. coli* leads to a dramatic increase in mortality where serious infections (blood stream) are caused by strains that are resistant to first line antibiotics, i.e., an increase in mortality from 18 to 40% (Foxman 2002). In the UK and across most of Europe *E. coli* is the main cause of Gram-negative blood stream infection (sepsis). Moreover, most of these bloodstream infections come from prior urinary tract infection (**Figure 1.1**). ExPEC disease is a term that is used for any *E. coli* infection that is found in the human body other than the gut.

The main reason why *E. coli* causes more ExPEC disease than any other human pathogen is closely related to the fact of its high carriage rate in human populations irrespective of geographic location or temperature. For example, *K. pneumonia*e is carried by about 20% of the human population and *Pseudomonas aeruginosa* by less than 1%. In fact, most *E. coli* infections are endemic arising from organisms carried within the patients own gut rather than being acquired from environment or

through food or drinking water. This is the major reason why *E. coli* is the main cause of serious infection (bloodstream infection) throughout Europe.

Since the beginning of the mandatory monitoring in the UK of *E. coli* bacteraemia in July 2011 until the beginning of the COVID-19 pandemic, the incidence rate of all reported cases of *E. coli* bacteraemia has risen yearly. This trend continued until the beginning of the COVID-19 pandemic. This rise was mostly driven by community-onset cases. After the pandemic, both the number of total reported cases and the incidence rates of community-onset cases went down, although the overall number of cases is still higher than what was seen when *E. coli* surveillance first began and has steadily risen again since the pandemic. In contrast, the incidence rate of hospital-onset cases hardly changed during the same period, even though a significant drop had been seen in the previous quarter (**Figure 1.1**) adapted from (Public Health England, 2021).

As mentioned before *E. coli* is one of the earliest bacterial species to colonise the gut during infancy, reaching extremely high densities (more than 10^9 cfu per gramme of faeces (Mitsuoka and Hayakawa 1973; Penders et al. 2006), before the proliferation of anaerobes (Syed et al. 1970). The density stabilises at roughly 10^8 cfu per gram of faeces after 2 years and stays there until it progressively declines in the elderly (Mitsuoka et al. 1975). However, the typical *E. coli* microbiota does assist its host in certain ways, primarily by inhibiting colonisation by pathogens (that is, by inducing colonisation resistance in the host) via the generation of bacteriocins and other processes (Rastegar Lari et al. 1990; Schamberger et al. 2004).

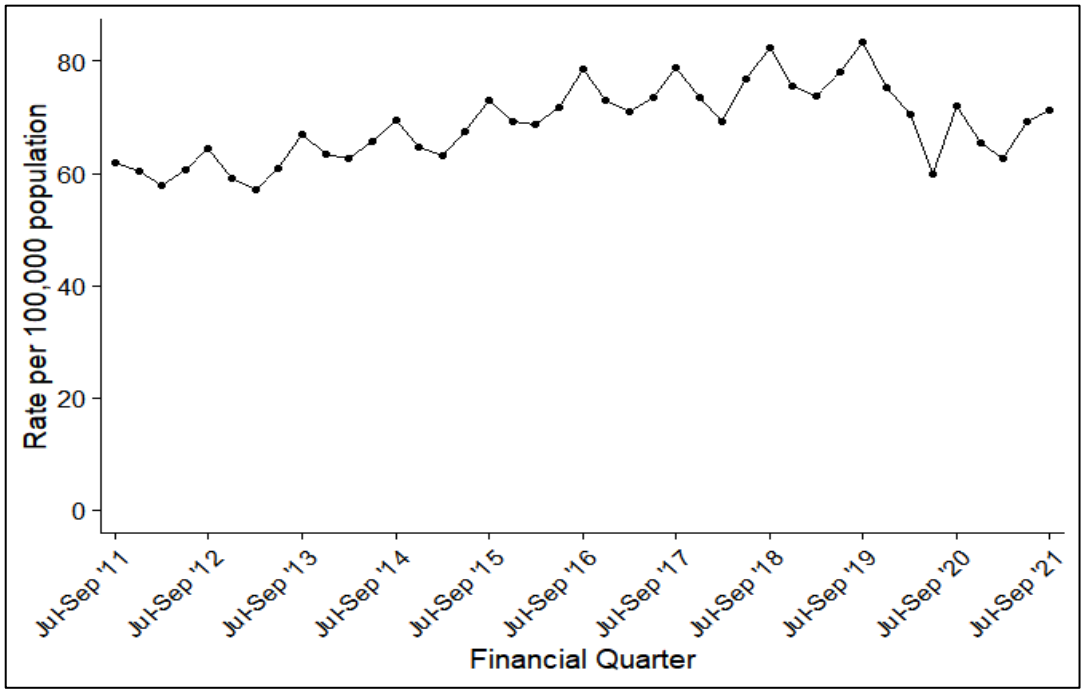


Figure 1.1 *E. coli* bacteraemia rate per 100,000 population (England, Wales, and Northern Ireland): 2011-2021

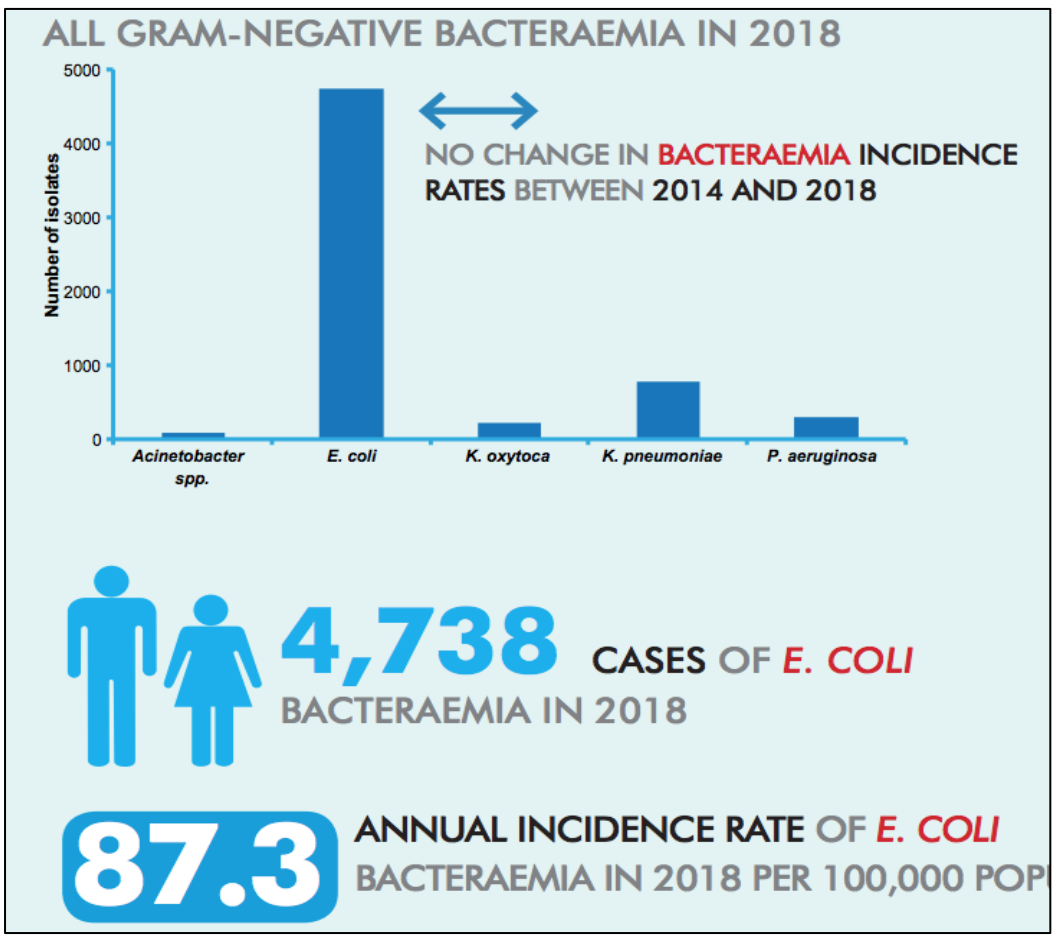


Figure 1.2 Result of incidence and Gram-negative bacterial cause of sepsis recorded in Scotland, the UK in 2018

Note that there is no recorded increase in sepsis in Scotland 2014-2018 as compared to the UK, but that the overall incidence is higher in Scotland than the whole of the UK.

1.6 *E. coli* genetic diversity

E. coli is one of the most genetically diverse bacterial species. Studies by the Ussery group in 2010 and in 2015 (Lukjancenko et al. 2010) (Land et al. 2015) give us interesting information on the genetic diversity found amongst individual *E. coli* isolates. They did a systematic pairwise comparisons of the genes found in fully sequenced *E. coli* strains available in the genetic databases in 2010 and 2015, respectively (**Figure 1.3 & 1.4**). As more *E. coli* whole genomes were added to the comparison, the number of unique genes (shown in blue in **Figure 1.3 & 1.4**) increased to 16,000 in the 2010 study and to c. 90,000 in the 2015 study. However, the number of genes found in all members of the *E. coli* species reduced drastically in the 2010 study to about 1,000 genes. As more *E. coli* genomes and closely related species were compared in the later 2015 study, the number of shared genes (the core *E. coli* genome) remained at about 1,000 genes. This essentially means that *E. coli* as a species easily loses and gains genetic information (Lukjancenko et al. 2010). About a third of *E. coli* genes are singletons, which means they only exist in one genome. A lot of these genomes are just rough draughts, and it is likely that gene fragments and errors in calling genes led to too many unique genes being predicted. Even with this in mind, there are probably more than 60,000 different gene families in *E. coli*, which was a lot at the time for a bacterial species with only about 2,000 sequenced members in 2015 (Land et al. 2015).

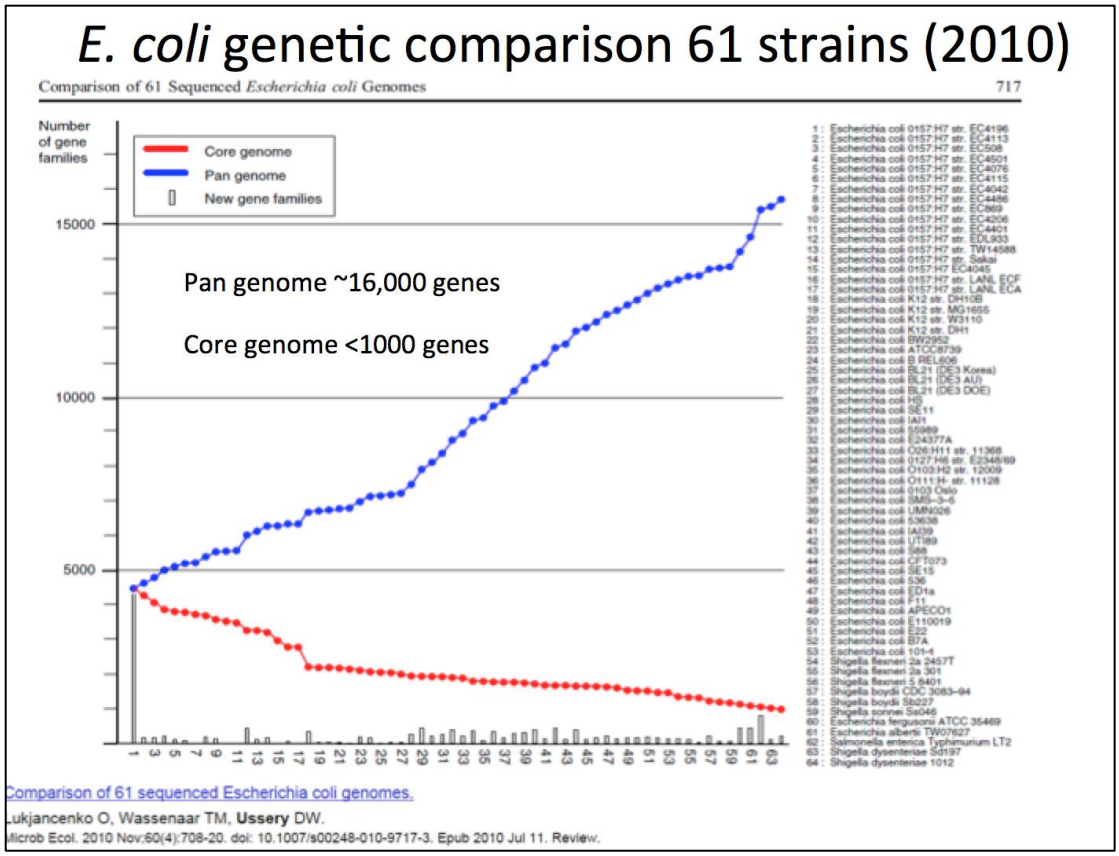


Figure 1.3 Comparison of the number of unique genes
(Pan-genome shown in blue) with number of shared genes (core genome shown in red) of a progressive pair-wise comparison of the 61 whole genome sequenced *E. coli* strains available in the genetic data bases in 2010 (Lukjancenko et al. 2010).

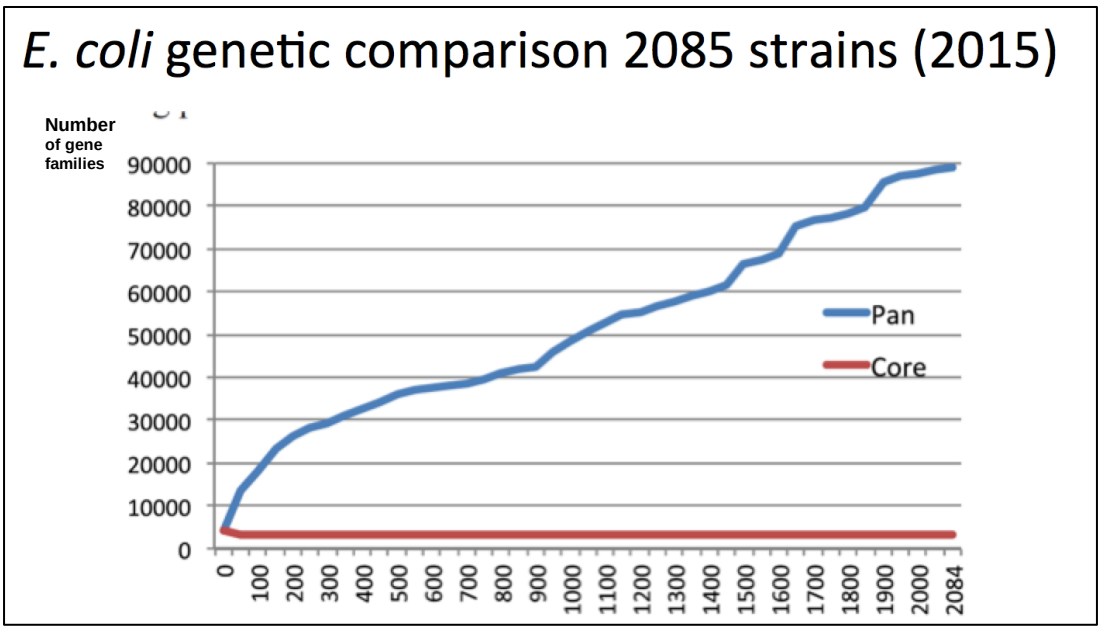


Figure 1.4 Comparison of the number of unique genes
(Pan-genome shown in blue) with number of shared genes (core genome shown in red) of a progressive pair-wise comparison of the 2085 whole genome sequenced *E. coli* strains available in the genetic data bases in 2015 (Land et al. 2015).

1.7 Antibiotics: discovery and development

Since ancient times, people have looked for medicines in the hope of relieving the symptoms of illness. Most of these treatments have included the application of salves or poultices to sick parts of the skin or body orifices (Chopra and Greenwood 2001). When used topically, honey is known for its healing properties and has a long history of usage for treating skin ulcers. Before the twentieth century, the only treatments that were scientifically proven to be effective for systemic infections were a few plant extracts that were active against parasitic worms or protozoa. The most notable of these was cinchona tree, which contains the alkaloids quinine and quinidine, which have been tremendously effective in preventing malaria since the seventeenth century. The only antibacterial chemicals that had even a relatively low benefit were mercury (for syphilis), chaulmoogra oil (for leprosy), and, since 1895, hexamine (methenamine) for cystitis. Other antibacterial drugs offered no effect at all. In 1882, Paul Ehrlich discovered a way to stain the tubercle bacillus. He envisioned that a toxic dye that would stain bacterial cells, but not human ones would likely kill bacterial cells whilst leaving human cells unharmed and devised the incredible idea of selective toxicity. Driven by this idea he used trypan red which stained the African trypanosome (the cause of sleeping sickness) to treat sleeping sickness. Further focussing on toxic dyes to treat syphilis. He discovered the world's first antibiotic Salvarsan. After hundreds of experiments on rabbits using arsenical drugs to cure syphilis, it was shown that salvarsan was effective. Ehrlich named it "Salvarsan" and introduced it to the public. Further research revealed that the 914th arsenical compound, named "Neosalvarsan," was easier to make and administer due to its higher solubility and greater curative impact. Salvarsan or Neosalvarsan

treated human syphilis (Piro et al. 2008). Paul Ehrlich, a German Jew, steeped in the knowledge of the Old Testament of the Bible and the knowledge of the teaching of clean and unclean related to infectious disease, named this new therapy “therapia sanitans magna”, which means the great sterilising principle or the therapy that makes clean. This amazing discovery was inspired and the beginning of the antibiotic era ushering in a new era in the effective treatment of diseases (Chopra and Greenwood 2001).

1.7.1 Sulfonamides

Following the groundbreaking research conducted by Paul Ehrlich (1854–1915) on the differential affinity of dyes for tissues and bacteria, the German dyestuffs industry was looking to use a series of dyes as antimicrobial agents. Prontosil was one of such dyes and the first member of the group known as the sulphonamides. Gerhard Domagk discovered an astounding function of the dye when he used it to protect mice from an infection with haemolytic streptococci, which would have otherwise been fatal and was awarded the Nobel Prize for his discovery in 1939. Being a Jew, he was held under house arrest in Germany by Adolf Hitler for the duration of the Second World War only to receive the prize after the war (Chopra and Greenwood 2001). This power of prontosil was rapidly proven in the clinic, most notably by Leonard Colebrook (1883–1967), who showed its usefulness in young women with life-threatening childbed fever.

Since 1939, many sulfanilamide derivatives have been made and caused a 50% reduction in soldiers dying from battle-field injuries due to infections between the first and second world wars. Many other antibacterial chemicals come out of the discovery of the sulphonamides. These included the antituberculosis drugs p-

aminosalicylic acid (1946), isoniazid (1952), and pyrazinamide (1952), as well as the first nitrofurans (1944), nalidixic acid (1962, the first quinolone drug), and trimethoprim (1967). But it was the discovery that natural products, “antibiotics" in the true sense of the word, were very effective at killing bacteria that sealed the deal on the therapeutic revolution in the treatment of infections (Chopra and Greenwood 2001). This momentous discovery and name originated from Selman Wacsman, a Ukrainian immigrant of Jewish descent who worked as a professor of microbiology of the soil and was the discoverer of Streptomycin and around another 20 antibiotics derived from soil organism. Again, this Jewish gentleman was driven by the teachings of the Old Testament of the bible and started his Nobel prize lecture with a quote from the book of Ecclesiasticus “The Lord has created medicines out of the earth; and he that is wise will not abhor them. “Ecclesiasticus, XXXVIII, 4. Available at (Waksman S 1952). He also has a quote from the Old Testament of the bible written on his gravestone from the prophet Isaiah that prophesised the discovery of antibiotics from the soil: Isaiah 45, 8 “The earth will open and bring forth salvation”.

1.7.2 Penicillins and β-lactam antibiotics

Howard Florey and Ernest Chain (A German Jew, escaping Nazi Germany) were examining the properties of naturally occurring bacteriolytic substances in the same year that Dogmark was first awarded the Nobel Prize (1939). Among these were two compounds, lysozyme and penicillin, which had been discovered by Alexander Fleming. Florey and Chain also examined the properties of other naturally occurring bacteriolytic substances (1881–1955). In 1929, when Fleming published his description of penicillin, which was the result of a mould that was a contaminant, he had just a passing interest in the idea that it may be used therapeutically. However, Chain had the ability to purify the active substance in 1940 and showed its amazing

efficacy in a speculative experiment with just eight mice infected with *Streptococcus pyogenes* with untreated animals serving as controls. After that, in 1941, the results of the successful human experiments were reported. The penicillin nucleus, also known as 6-aminopenicillanic acid, was found in fermentation products by researchers at Beecham Research Laboratories in 1959. This discovery paved the way for the invention of methicillin, cloxacillin, ampicillin, and the other semi-synthetic penicillins (Chopra and Greenwood 2001).

Cephalosporins are produced by a mould and are known to produce cephalosporin antibiotics which was discovered by Guiseppe Brotzu, a public health official from Sardinia. He discovered this from extracts of sea water near a sewage outlet in Sardinia after being perplexed by the observation that children swimming near a sewer outlet during a typhoid outbreak appeared to never develop disease (Muñiz et al. 2007). Through further separation of the antibiotic complex in the research labs in Oxford, the precursor to the cephalosporins, known as cephalosporin C, was shown to be present as a small component. Similar to the development of penicillins, semisynthetic derivatives were created, which eventually resulted in the creation of a broad family of chemicals that are currently the most used β-lactam antibiotics. In subsequent years, it became apparent that naturally occurring antibacterial chemicals that contain a β-lactam ring are widely dispersed throughout the environment. In later years, a few antibacterial agents, such as carbapenems, monobactams, and the b-lactamase inhibitor clavulanic acid, were developed for clinical use (Chopra and Greenwood 2001).

1.7.3 Other antibiotics

Chloramphenicol (1947), chlortetracycline (1948), erythromycin (1952), vancomycin (1956), lincomycin (1962), and gentamicin (1963) are just some of the antibiotics that

were discovered because of the pharmaceutical industry's massive screening programmes in response to the discovery of streptomycin from the soil (Toleman 2017) (**Figure 1.5**).

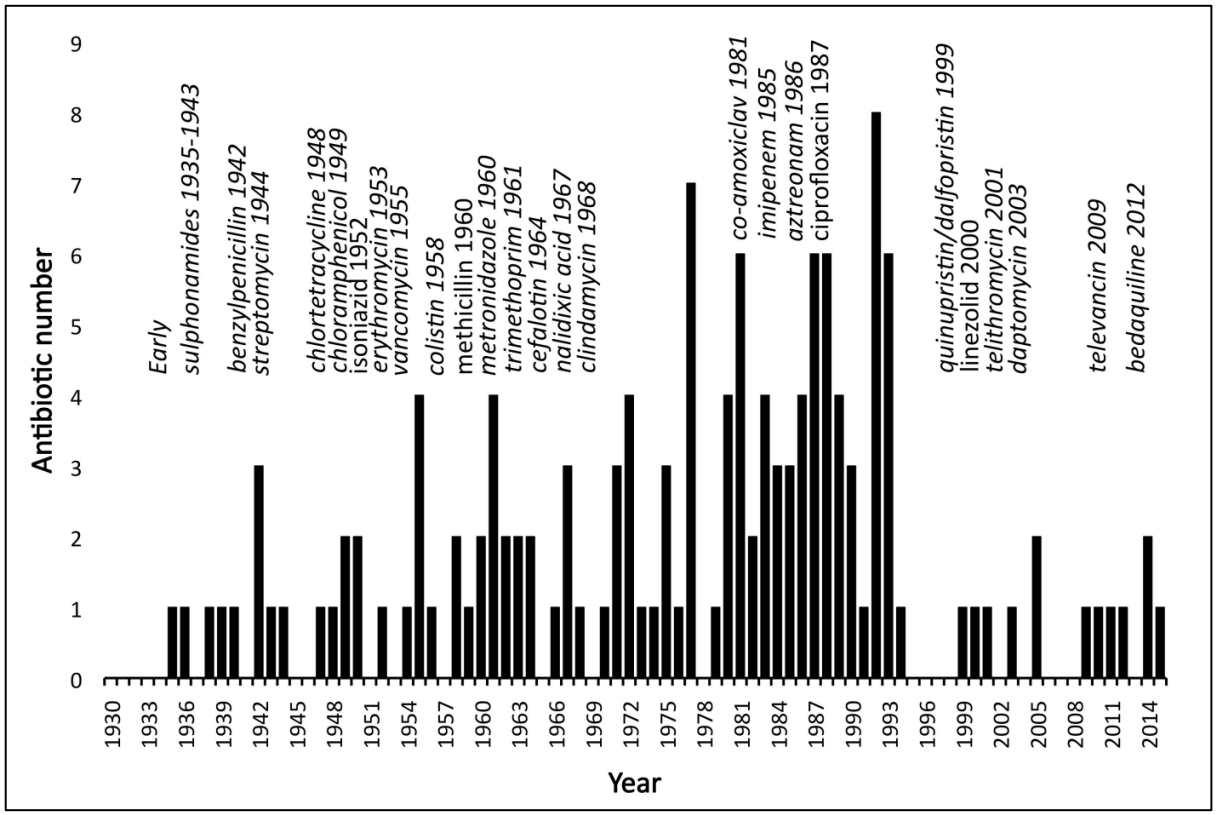


Figure 1.5 Timeline of antibiotics discovery
The graph displays the number of antibiotics released onto the pharmaceutical market each year since 1935 (Toleman 2017).

Following the discovery of molecular targets within bacterial cells, reasonable screening techniques were developed, which led to the development of efficient antibiotics with a variety of mechanisms of action (Toleman 2017).

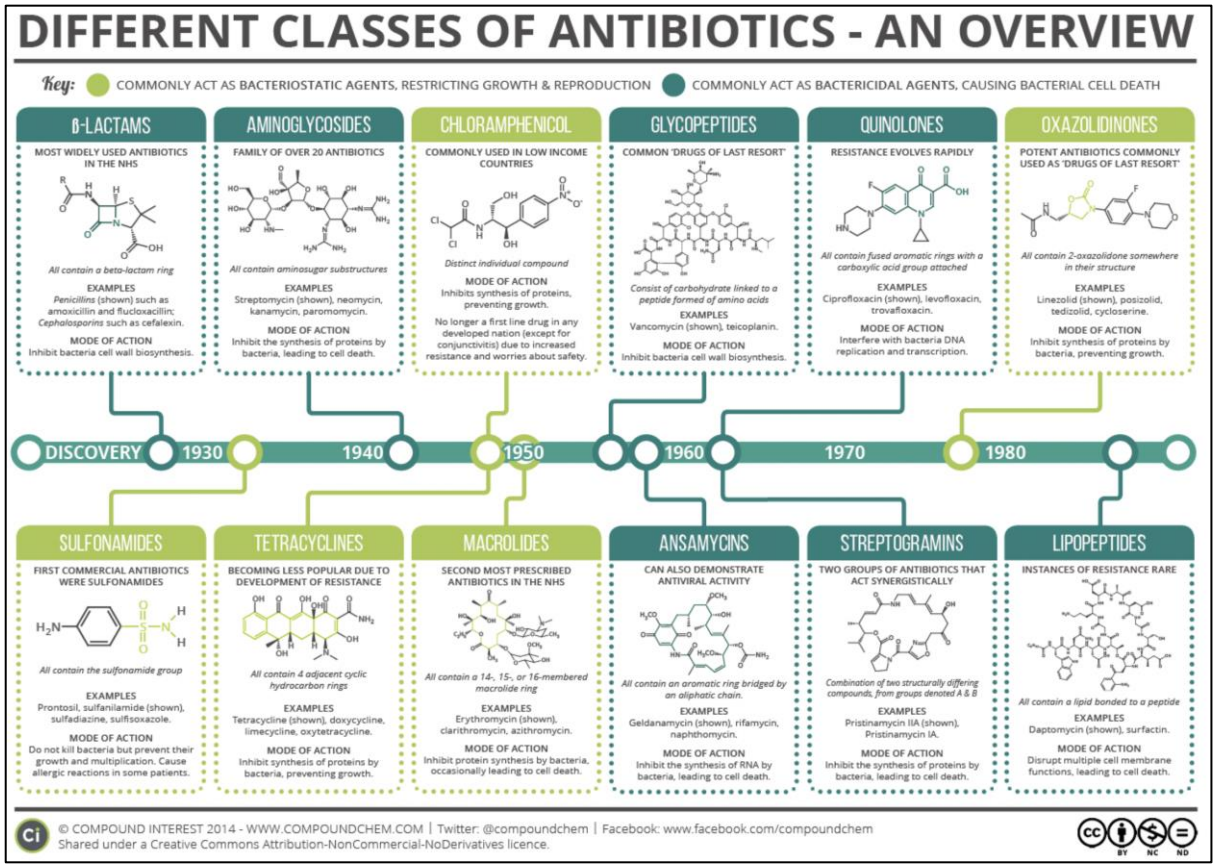


Figure 1.6 A brief overview of different classes of antibiotics
Taken from compound interst website WWW.COMPOUNDCHEM.COM

1.8 Mode of actions

The term "selective toxicity" originally envisioned by Ehrlich refers to the goal of all antimicrobial chemotherapy, which is to suppress or kill the organism that is causing the infection without causing harm to the host. To achieve their selective effect, most antibacterial agents are able to take advantage of differences in the structure or metabolism of bacterial and mammalian cells and the undesirable side effects of antibacterial therapy are typically the result of idiosyncratic effects that are unrelated to the primary mechanism by which antibacterial agents work (Valent et al. 2016).

1.8.1 Inhibitors of cell wall synthesis

There are several points along the process of bacterial cell wall formation that serve as potential therapeutic targets for antibacterial medicines. The first step is the creation of N-acetyl muramic acid (NAMA), which occurs as the result of the condensation of phosphoenolpyruvate with N-acetylglucosamine (NAG). This reaction is one that may be blocked by the phosphonic acid antibiotic such as fosfomycin. Next, NAMA is replaced by five different amino acids, of which the very last two, D-alanyl-D-alanine (D-ala-d-ala), are added as a unit. D-alanine is produced from L-alanine by the activity of alanine racemase, and the dipeptide is produced through the action of a ligase. Cycloserine, a second line antituberculosis agent, acts as a competitive inhibitor for both events. Afterwards, any amino acids that are required for interpeptide bridges are added at this point, and the cell wall unit is completed when NAG is linked to the NAMA-peptide (**Figure 1.7**) (Chopra and Greenwood 2001).

All this activity takes place within the cytoplasm of the cell, and the unit of the cell wall must then be transferred across the cytoplasmic membrane to the location where the peptidoglycan will develop. Lipid carrier molecules in the membrane, (55-carbon isoprenyl phosphate), are responsible for transporting the molecule across the membrane. During this stage of the process, the lipid will pick up an extra phosphate group, which will later be eliminated to restore its carrier function. Bacitracin, a cyclic peptide antibiotic that is too toxic for systemic usage but is utilised in topical treatments, inhibits this dephosphorylation process (**Figure 1.7**) (Chopra and Greenwood 2001).

The glycopeptides antibiotics, such as vancomycin and teicoplanin, block the cell wall cross-linking by binding to the terminal Dala-D-alanine residue of the cross-link.

The transpeptidation process is also the site of action by the β-lactam antibiotics including penicillins, cephalosporins, carbapenems and monobactam aztreonam. This achieved by a slightly different methodology to the glycopeptides. The β-lactam inhibits the actual enzymes (penicillin-binding proteins) by binding to their active sites (Toleman 2017).

1.8.2 Inhibitors of bacterial protein synthesis

Protein synthesis can be inhibited by a variety of clinically relevant antibacterial drugs. This is accomplished by directly limiting the actions of ribosomes, which are the cellular organelles that are responsible for decoding the genetic instructions to generate proteins. Antibacterial protein synthesis inhibitors exert their unique activity due to structural variations between bacterial and mammalian ribosomes (Etebu and Arikekpar 2016). Antibacterial protein synthesis inhibitors have a specific action because of structural differences between bacterial and mammalian ribosomes. These structural differences allow antibacterial agents to selectively exploit unique binding sites in bacterial ribosomes, which results in the inhibition of bacterial protein synthesis. The production of bacterial proteins is a complicated process that requires the generation of active 70S particles that are capable of catalysing peptide bonds between incoming amino acids during a process that is known as the elongation cycle (**Figure 1.7**). This cycle is either inhibited or altered by most of the clinically relevant inhibitors of bacterial protein synthesis, such as lincosamides, macrolides, tetracyclines, chloramphenicol, aminoglycosides, streptogramins, and fusidic acid. Linezolid, which is a member of a new class of antibacterial agents called the oxazolidinones, prevents the formation of the 70S initiation complex by binding to the 50S ribosomal subunit. This contrasts with the majority of protein synthesis

inhibitors, which have an effect on the elongation cycle (Chopra and Greenwood 2001).

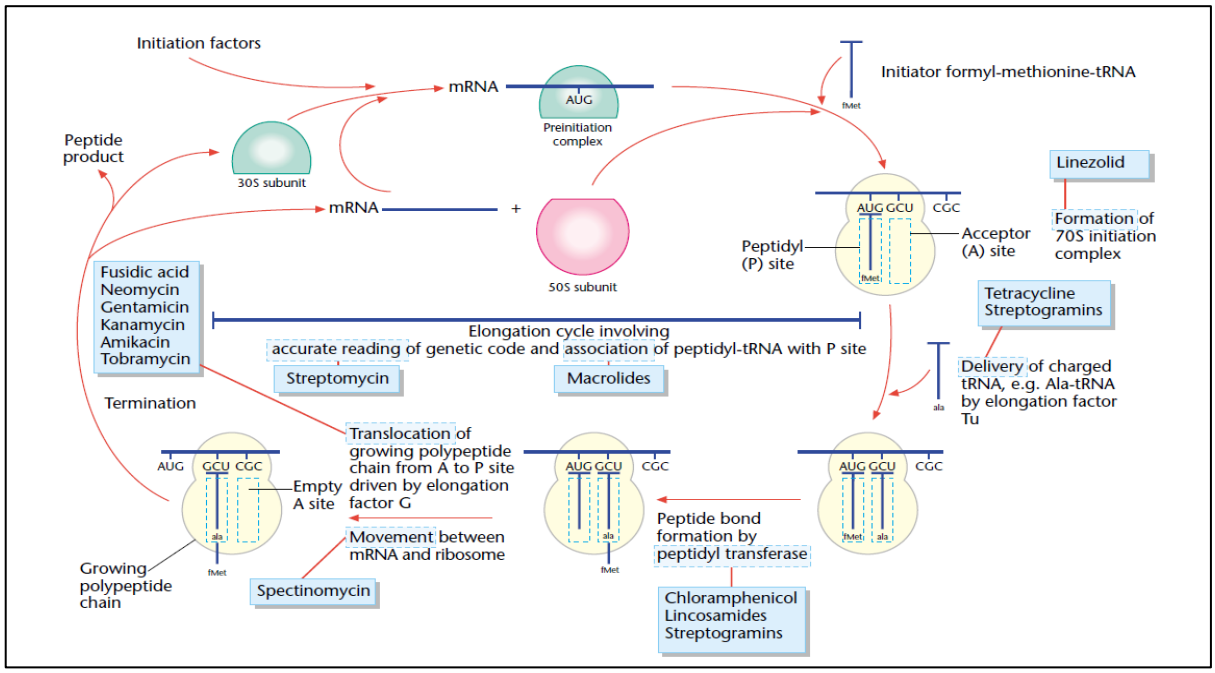


Figure 1.7 Bacterial protien synthesis, showing the steps that are inhibited by various antibacterial agents (Chopra and Greenwood 2001)

1.8.3 Inhibitors of nucleic acid synthesis

Antibacterial drugs can specifically interfere with the processes of RNA and DNA synthesis, both of which are necessary for bacterial cell survival (Kohanski et al. 2010).

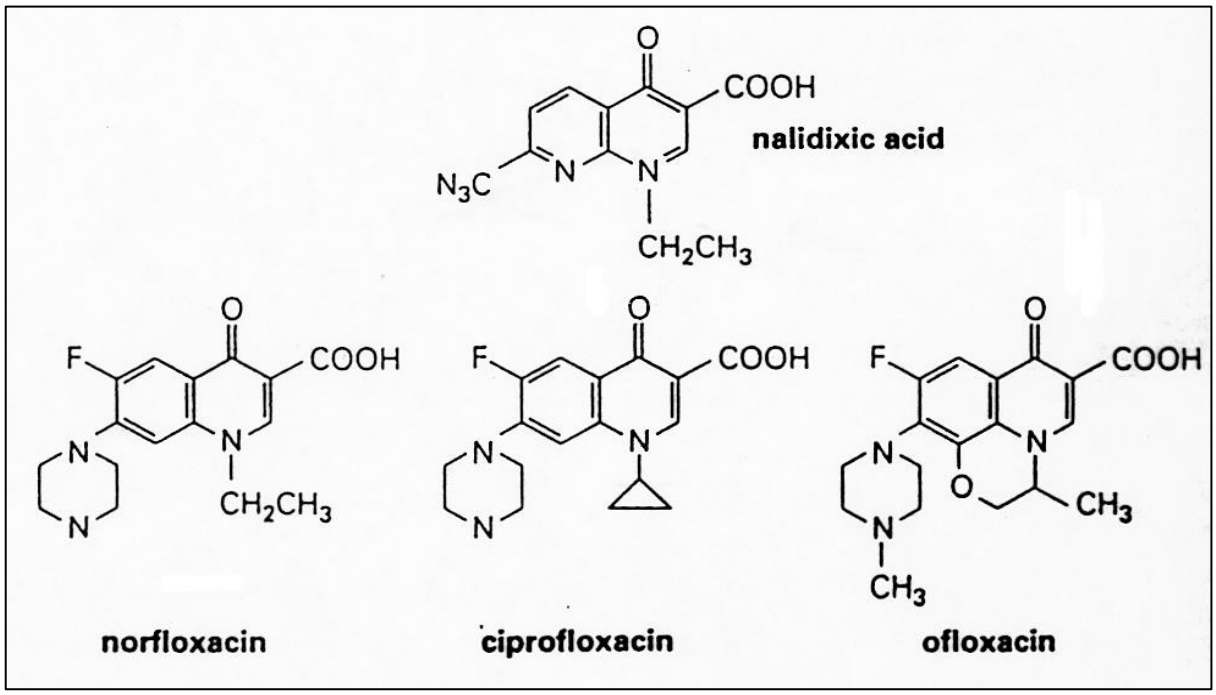


Figure 1.8 Inhibitor of nuclic acid synthesis. Quinolones: ciprofloxicin, gatifloxicin, moxifloxicin (inhibitors of DNA gyrase and topoisomerase IV)

1.8.4 Inhibitors of RNA synthesis

RNA polymerases are responsible for mediating the transcription of structural genes. These enzymes do this by catalysing the beginning of the formation of RNA molecules on a DNA template and continuing their growth. Rifampicin is an antibiotic that belongs to the rifamycin group that binds specifically to the β-subunit of RNA polymerase. As a result, it inhibits the function of the holoenzyme (Es) that begins the process of RNA synthesis (Toleman 2017). Inhibition most likely arises from interference with the production of the first phosphodiester bond in the RNA chain. The binding of RNA polymerase to the DNA template is not prevented. Rifampicins selective antibacterial effect can be explained by the structural variations that exist between the RNA polymerases found in bacterial cells and those found in mammalian cells. Mammalian RNA polymerases, in contrast to bacterial enzymes,

have 10 or more polypeptide chains, none of which contain sites for the binding of rifampicin. Bacterial enzymes are found in cells (Kohanski et al. 2010).

1.8.5 Inhibitors of DNA synthesis

The quinolone group of antibiotics function by inhibiting the packaging of the bacterial DNA, which is necessary to fit this huge molecule within the bacterial cell and to allow access to the proteins necessary for replication and expression of the DNA. The DNA is packaged by supercoiling that involves topoisomerase enzymes (Karl et al. 2008). Positively supercoiled DNA is accomplished by topoisomerase IV, and it does this by nicking a single strand of the DNA twisting clockwise and resealing the gap. DNA gyrase, which is also known as topoisomerase type II, twists the DNA in an anticlockwise direction and forms a double stranded break in the DNA and then reseals after twisting. Antibacterial drugs belonging to the quinolone class, such as nalidixic acid and the more modern fluoroquinolones, can specifically inhibit DNA gyrase and topoisomerase IV, and as a result, they are able to halt the necessary process of bacterial DNA replication. DNA gyrase is the principal target of quinolones when dealing with Gram-negative bacteria; however, DNA topisomerase IV is the key target when dealing with Gram-positive bacteria. These medications do not bind directly to DNA gyrase or topoisomerase IV; rather, they generate stable complexes with DNA and enzymes that inhibit the topoisomerases ability to carry out their normal functions (Chopra and Greenwood 2001) (**Figure 1.9**).

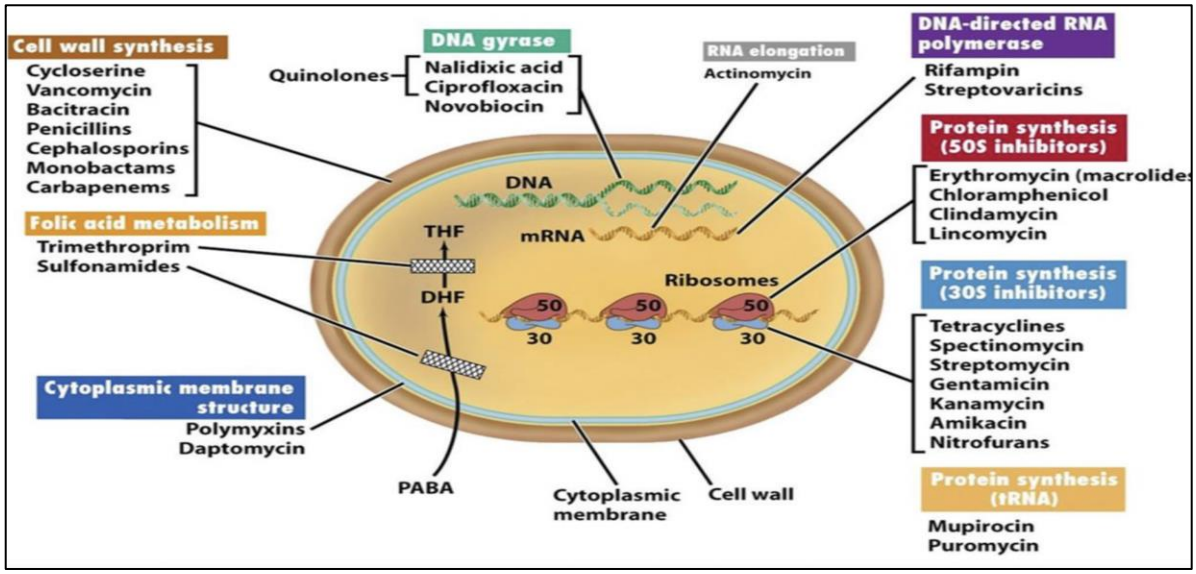


Figure 1.9 Antibiotics target sites (Etebu and Arikekpar 2016)

1.8.6 Mode of action of β-lactam

A variety of distinct types of wall growth are required for the growth and division of the bacterial cell. To carry out these coordinated activities, a battery of enzymes with transpeptidase, endopeptidase, and carboxypeptidase activity is required. The β-lactam antibiotics can interact differentially with several of the "penicillin-binding proteins" (PBPs). There are at least seven of these PBPs found in *E. coli*, and they are listed in the order of decreasing molecular weight. The killing of bacteria by β-lactam antibiotics requires the participation of three different proteins: the PBP-1, PBP-2, and PBP-3. Even though they bind at varying degrees of affinity, most β-lactam drugs can bind to all of these PBPs. They cause lysis of susceptible Gram-negative bacteria when present at sufficient concentrations. This occurs because the faulty cell wall is unable to shield the cell from osmotic rupture (Kapoor et al. 2017). Other β-lactams, such as cephalexin, cephradine, and aztreonam, specifically bind to PBP-3, a protein that is necessary for the formation of septum and enable the

cells to continue growing into long multinucleate filaments before death occurs. Mecillinam, imipenem, and clavulanic acid, which is an inhibitor of β-lactamase, are some of the other antibiotics that preferentially bind to PBP-2. Affected cells progressively round up before being destroyed by osmotic lysis because this PBP is employed in the elongation of the cell wall cylinder (Chopra and Greenwood 2001).

1.9 The β-lactamases

These enzymes can break β-lactam molecules, making them inert. As a result, they can enable β-lactam resistance, either on their own or in conjunction with other factors. It is the most prevalent resistance mechanism that leads to the broad prevalence of resistance among Gram-negative microorganisms (Bush and Jacoby 2010).

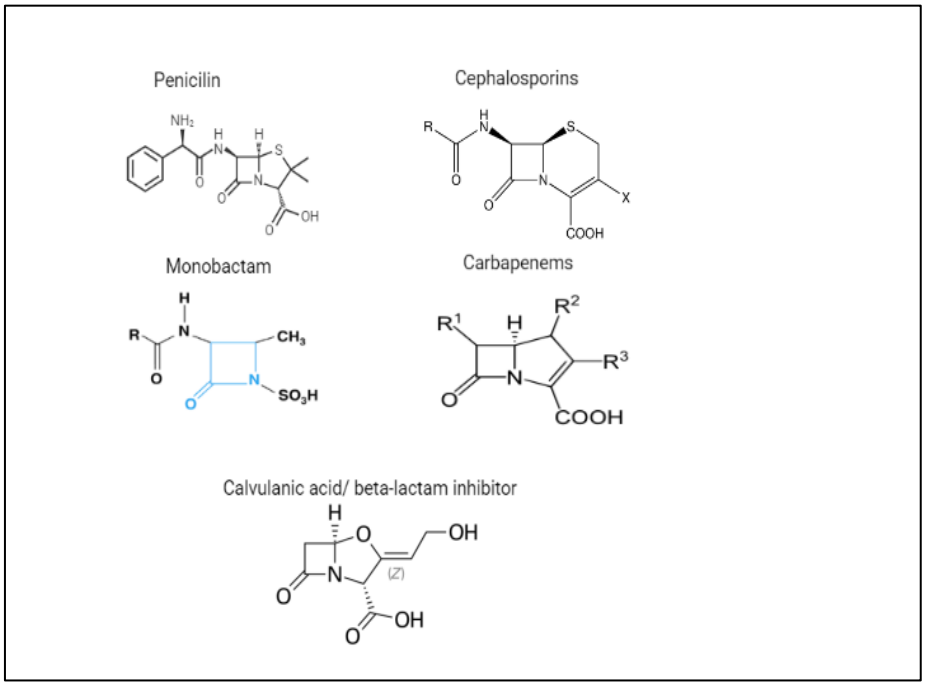


Figure 1.10 The structure of β-lactam compounds

1.9.1 Classification of β-lactamases

Two different classification schemes are used for these enzymes. The first is the Ambler molecular classification, also just known as a molecular classification. This classification is based on the protein sequence, and it divides β-lactamases into four different molecular classes: A, B, C, and D (Ambler 1980). Each of these classes is distinguished from the others by a distinct and conserved pattern of amino acids, especially in the active site of the enzyme. Enzymes belonging to classes A, C, and D hydrolyse their substrates by producing an acyl enzyme through an active site serine amino acid. In contrast, β-lactamases belonging to class B are metalloenzymes that use at least one active-site zinc ion to assist β-lactam hydrolysis. The molecular classification thus gives some understanding of enzyme function, i.e., by offering class-specific resistance to various β-lactam antibiotics.

Ambler initially classified β-lactamases into two main categories: class A, which includes active-site serine β-lactamases, and class B, which includes metallo-β-lactamases that are only active in the presence of a bivalent metal ion, usually Zn^{2+}. Later, a novel group of serine β-lactamases was discovered. These enzymes had very little sequence resemblance to the previous class A enzymes that were known at the time and were therefore named class C. Class C β-lactamases are sometimes referred to as AmpC β-lactamases. A further group of serine β-lactamases was also subsequently found mostly referred to as the OXA β-lactamases. Since this group had no similarity to either class A or class C, they were designated as class D. However, there is still enough structural similarity between the three classes of serine β-lactamases to make it abundantly clear that they are closely related and perhaps related to one another through descent from a common ancestor. This is even though the three classes of serine β-lactamases are distinct enough to prevent alignment programmes like BLAST from finding any detectable sequence similarities between them.

The second category is the Bush, Jacoby, Medeiros functional classification that classifies β-lactamases by substrate and inhibitor profiles and is a more clinically based classification system that was originally developed in 1989. Based on their substrates and inhibitors, β-lactamase enzymes can be put into three main groups: Group 1 are cephalosporinases, which are not well inhibited by clavulanic acid; Group 2 are penicillinases, cephalosporinases, and broad-spectrum β-lactamases, which are generally inhibited by active site-directed β-lactamase inhibitors; and Group 3 are metallo β-lactamases, which break down penicillins, cephalosporins, and carbapenems and that practically all inhibitors have no or little effect against (Bush et al. 1995).

This classification system was further updated with the identification of four β-lactamases grouping (Bush and Jacoby 2010):

- Group 1 are cephalosporinases, which are poorly inhibited by clavulanic acid.
- Group 2 are β-lactamases, which are typically inhibited by active site-directed β-lactamase inhibitors, and which belong to molecular classes A or D.
- Group 3 are metallo-β-lactamases, which are only moderately inhibited by all classical β-lactamase inhibitors apart from the non-clinically useful inhibitors EDTA and p-chloromercuribenzoate (pCMB).
- Group 4 are penicillinases that are not inhibited by clavulanic acid.

The 1989 classification key groupings were mostly preserved. However, three alterations were included. Firstly, it was chosen to categorise derivatives of enzymes in groups that retain the "2b" prefix with the 2be classification because of the rise of TEM- and SHV-derived β-lactamases that have extended-spectrum β-lactamases activity to demonstrate that these enzymes are descendent from group 2b enzymes and have an extended spectrum of activity. Similarly, a new group 2br has been created for structurally related group 2b β-lactamases with lower affinity for β-lactamase inhibitors. It is predicted that similarly named β-lactamases created from enzymes belonging to other groups may one day be used to identify closely related ones.

The group 2f β-lactamases, which hydrolyse carbapenems and are only moderately inhibited by clavulanic acid, are the third group of enzymes to be introduced to the scheme. It is now known that these enzymes have a serine in their active site. More novel enzymes that have been characterised in 1989 to 1995 are among the extensive additions to the 1989 tables. A few older enzymes were also re-evaluated

utilising inhibitors or substrates that were not around when the first results for those β-lactamases were reported (Bush and Jacoby 2010).

Representative β-lactamases from all molecular classes are divided into categories mostly based on published functional properties. Group 3 was given to β-lactamases that were inhibited by EDTA, a group which at the time had only a very small number of β-lactamases but which has expanded significantly in the last couple of decades. Following the separation of the metalloenzymes from the other β-lactamases, the enzymes were classified based on the substrate profiles. The following factors were taken into account when establishing the priority order. First, it was determined if an enzyme would be categorised as a penicillinase or a cephalosporinase by comparing the relative hydrolysis rates for benzylpenicillin and cephaloridine. An enzyme was then classified as either a penicillinase or a cephalosporinase if it hydrolysed one of these substrates at a relative rate that was roughly 30% lower than that seen for the other β-lactam. It should be noted that some cephalosporinases broke down benzylpenicillin but not other penicillins. Based on this activity and the different ways the microbe that made the enzyme reacted to penicillins and cephalosporins, it was decided that the organism belonged in group 1. Broad-spectrum enzymes were the ones that broke down both substrates at about the same rate. By looking at how fast carbenicillin or cloxacillin (oxacillin) was broken down by penicillinases, more subgroups of enzymes were found. If an enzyme broke down cloxacillin or oxacillin at a rate >50% slower than benzylpenicillin, it was put in group 2d. This group may also contain enzymes that break down carbenicillin. Most of the time, clavulanic acid does not stop these enzymes as well as it stops most group 2 β-lactamases. If carbenicillin was broken down at a certain rate, then the enzyme was put in group 2c because it broke down benzylpenicillin at a rate of >60% that for benzylpenicillin and

cloxacillin or oxacillin at a rate <50% that for benzylpenicillin. If the rates of hydrolysis for ceftazidime, cefotaxime, or aztreonam, which are extended-spectrum β-lactam antibiotics, were higher than 10% that for benzylpenicillin, the enzyme was put in group 2be, which is made up of extended-spectrum β-lactamases. This group was first named "extended-broad-spectrum β-lactamases" because the enzymes in this class also broke down broad-spectrum antibiotics like penicillin and cephalosporins. Group 2e was made up of cephalosporinases that broke down cefotaxime well but did not break down penicillin well and were prevented by clavulanic acid. There were also several exceptions taken into consideration to the 2be category. Even though the hydrolysis criteria were not met stringently, large increases in hydrolysis rates for ceftazidime were noted when compared with those of the parent enzymes. This resulted in increased MICs of that cephalosporin for TEM-producing organisms. The decision was made to include β-lactamases such as TEM-7 and TEM-12 enzymes that were derived as a result of point mutations in the TEM-2 and TEM-1 genes, respectively (Bush and Jacoby 2010).

1.10 Important β-lactamase enzymes

1.10.1 Extended-spectrum β-lactamases (ESBLs)

ESBLs are mostly found on plasmids and are known for their ability to break down oxyimino-cephalosporin (3rd and 4th-generation cephalosporins) and monobactams, but not cephamycins like cefoxitin and carbapenems like meropenem, imipenem, ertapenem, and doripenem. In addition to this, they are typically susceptible to β-lactamase inhibitors such clavulanic acid, sulbactam, and tazobactam (Batchelor et al. 2005). ESBLs are traditionally defined as enzymes that were originally derived by mutation from a narrow spectrum parent ESBL enzyme such as TEM and SHV. As a

result, these enzymes gained the ability to inactivate broad spectrum cephalosporins, penicillins, and aztreonam, but not cephamycins (cefoxitin) or the carbapenems. Additionally, these enzymes can be inhibited by β-lactamase inhibitors, such as clavulanic acid (Ur Rahman et al. 2018a). The CTX-M, TEM, and SHV kinds of ESBLs have been shown to be the most successful in terms of promiscuity and dispersion over a wide range of epidemiological niches and have been reported in a wide range of pathogens (Ur Rahman et al. 2018).

1.10.2 CTX-M

CTX-M-1 and CTX-M-2 β-lactamases were the first CTX-M-like enzymes to be discovered (Rossolini et al. 2008). Since 2000, *E. coli* producing CTX-M enzymes have emerged worldwide as important causes of community-onset urinary tract infections (UTIs). This phenomenon has been called the CTX-M pandemic (Cantón and Coque 2006; Pitout and Laupland 2008), and there are now >200 variants of CTX-M (see (Beta-Lactamase DataBase (BLDB): Structure and Function)). CTX-M enzymes are the most common type of ESBLs found in most area of the world (Rossolini et al. 2008). The CTX-M-15 enzyme is the most common variant found worldwide and has been increasingly described in community isolates, particularly associated with infections in healthcare-associated patients (Cantón and Coque 2006).

1.10.3 SHV

SHV originates from the chromosome of *Klebsiella pneumonia*. It gives resistance to broad-spectrum penicillins like ampicillin, tigecycline, and piperacillin, but not to oxyimino substituted cephalosporins. Mutation to SHV-1 led to the evolution of the SHV family. Some of which confer an extended spectrum resistance. The evolution

of the SHV β-lactamase happened at the same time as the evolution of the TEM β-lactamases (Shaikh et al. 2015b).

1.10.4 TEM

The substrate and inhibitor profiles of TEM β-lactamases were like those of SHV-1 when they were initially identified in 1965. That is TEM-1 is capable of hydrolysing penicillins and first generation cephalosporins. However, since 1965, TEM β-lactamases have undergone mutations-producing variants that may have ESBL activity. TEM-3 is one of these varieties http://bldb.eu/M-BLDB.php?prot=TEM-1 (Shaikh et al. 2015).

1.11 Carbapenamases

Carbapenemases are the most flexible family of β-lactamases, with their wide spectrum of substrate specificity and are unsurpassed by any other β-lactam-hydrolysing enzyme. Although many of these enzymes are referred to as "carbapenemases", they recognise practically any β-lactams, and the majority are resistant to inhibition by all β-lactamase inhibitors that are available for treatment. The hydrolytic mechanism at the active site divides carbapenemases into two molecular families. The first was from Gram-positive bacilli that initially produced carbapenemases. These were metalloenzymes since EDTA inhibited them, unlike other β-lactamases. Subsequent research showed that all metallo-carbapenemases include at least one zinc atom at the active site to hydrolyse a bicyclic β-lactam ring. In the 1980s, another subset of carbapenem-hydrolysing enzymes was reported, but EDTA did not inhibit it. Later research demonstrated that these enzymes used serine at their active sites and were inactivated by β-lactamase inhibitors clavulanic acid and tazobactam (Queenan and Bush 2007).

All carbapenemases were thought to be species-specific β-lactamases encoded on chromosomes until the early 1990s. Formerly established patterns of carbapenemases spread have been disrupted by the discovery of plasmid-encoded metallo-β-lactamases in *Pseudomonas aeruginosa* (IMP-1), *Acinetobacter baumannii* (ARI-1, OXA-23), and *Klebsiella pneumoniae* (KPC-1). What was originally thought of as an issue of clonal expansion has transformed into a problem of interspecies dispersion that affects the entire world (Codjoe and Donkor 2017). Common carbapenemases that are now found worldwide include OXA-48, KPC and NDM. NDM is particularly prevalent in *E. coli* around the world and originated from S. Asia. This enzyme is the result of the very recent fusion of two different antibiotic resistance genes: an aminoglycoside resistance gene *aphA6* and a metallo-β-lactamase (Toleman et al. 2012).

1.12 Resistance in *E. coli*

Since the first use of antibiotics, bacterial resistance to these lifesaving drugs was quickly observed. Both Domagk and Alexander Flemming included observations of antibiotics resistance in their noble lectures (See (Gerhard Domagk – Facts - NobelPrize.org; Sir Alexander Fleming - Facts)).

Antibiotic or antimicrobial resistance develops when normally effective antibiotics or antimicrobials are no longer able to eradicate the pathogens. If bacteria can live at high dosage of permitted medications, then they are considered resistant, and resistance can be observed in any bacterium. When the antibiotic does not kill the pathogen, this is called drug tolerance or drug failure. When an organism develops resistance to more than one class of antibiotic, we refer to that organism as being

multidrug resistant. Over time, bacteria can acquire these traits to withstand a once-effective medication (Ur Rahman et al. 2018).

These phenomena of resistance emergence had expedited humankinds attempt to overcome the resistance phenomenon by discovering antibiotics of the following generation, such as cephalosporin. The search for new antibiotics began in order to find drugs that could kill all existing pathogens. This led to the discovery of 2nd-generation (cefoxitin, cefotetan, cefmetazole, cefaclor, cefpodoxime, and cefuroxime), 3rd-generation (cefxime, cefodizime, cefotaxime, cefteram, and ceftizoxime), and 4th-generation (cefepime, cefluprenam, cefoselis, cefozopran, cefpirome, and cefquinome) cephalosporin drugs; at the same time, microbes are battling one another for supremacy and searching for new characteristics to adopt.

In addition, several nations do not require prescriptions to access antibiotics, which contributes to the rise of antibiotic resistance (Martens and Demain 2017). Antibiotics that are now available can be classified as either natural, synthetic, or semisynthetic, which refers to a natural substance that has been altered, or synthetic, which refers to antibiotics that have been chemically developed in labs and include sulphonamides and quinolones (Normark and Normark 2002).

1.12.1 Antibiotic resistance gene mobilisation and horizontal gene transfer

A strain may acquire resistance from another species or genus. Plasmids and other mobile genetic elements carry most antimicrobial resistance genes, which can spread to other bacteria (von Wintersdorff et al. 2016). The benefits of antibiotics for the treatment of infectious illnesses have been taken for granted in the field of public health ever since the pioneering work on antibiotics carried out by scientists such as Alexander Fleming, Paul Ehrlich, Gerhard Domagk, and others. Domagk and

Flemming both observed resistance but thought that all would be mutation based and so thought it very unlikely that resistance to several antibiotics could occur at one point in time for an organism to survive the application of more than one antibiotic.

1.13 Resistance gene transmission

Since the beginning of the antibiotic era the most common form of antibiotic resistance has been due to horizontal gene transfer (HGT) between bacterial organisms. HGT involves the movement of resistance genes often from environmental bacteria into clinically relevant ones such as *E. coli.* HGT can be described by two processes that is movement within bacteria and movement of resistance genes between bacteria. It is only the combination of both movement events that makes it possible for a bacterium to become resistant by HGT. Movement within bacteria is accomplished by transposable elements often called transposons or jumping genes; whereas, movement between bacteria is mostly accomplished by conjugative plasmids and integrative and conjugative elements (ICE) (Toleman and Walsh 2011).

1.13.1 Movement within bacteria

As mentioned above movement within bacteria is accomplished by transposons. There are three different types of transposons, including Insertion sequences, composite transposons, and unit transposons. Insertion sequences are usually composed of a single gene and inverted repeats of a small number of nucleotides (10-20 bp) found at either end of the single gene. The gene encodes an enzyme capable of cutting the DNA at the ends of the element within the inverted repeat sequences and enabling movement from one position in the DNA to another, such

as from the chromosome to a resident plasmid or from the plasmid to the chromosome or from one position on a plasmid or chromosome to another position. Composite transposons are the second type of transposon and simply consist of two usually identical insertion sequences that flank a resistance gene or other types of genes. The two insertion sequences move in tandem and mobilise any DNA found between them. The last type of transposon is a unit transposon, and these are usually much more complex consisting of several genes involved in movement and a number of genes such as a mercury resistance operon or an integron (Toleman and Walsh 2011).

1.13.2 Movement between bacteria

DNA transfers from one bacterial cell to another through one of three ways: conjugation, transformation, or transduction. Transduction involves phage-mediated DNA transfer (Davies and Davies 2010). In contrast, transformation is described as the uptake of the exogenous DNA from the environment (Dubnau 1999). However, the most important method of movement between bacteria is conjugation, which is made possible by conjugative plasmids (Broda et al. 1979) or ICE (Salyers et al. 1995; Burrus et al. 2002). The discovery of conjugation was with the discovery of the misnamed sex factor F, which is in fact a conjugative plasmid (Broda et al. 1979). The power of conjugation lies in the fact that some plasmids termed promiscuous plasmids can transfer between different species of bacteria; however, transformation and transduction tend to be found mostly with transfer of genetic information between identical or very similar bacteria. Transposons cannot by themselves transfer between cells and can only do this by “piggybacking” on a conjugative plasmid or ICE.

1.13.3 The Class 1 integron

An integron is a genetic structure that has a place for site-specific recombination that may be utilised to capture additional DNA sequences, also known as gene cassettes. The integrase gene *int* gene, which is found next to the *attI* recombination site, is responsible for encoding a recombinase enzyme, which is also known as an integrase (Mötsch and Schmitt 1984). This enzyme can either incorporate a gene cassette into the *att*I site or remove a gene cassette at the *att1* site. The transposon *Tn5090* (also known as *Tn402*) captured the class 1 integron either before the antibiotic era or at the beginning of the antibiotic era and can mobilise the whole integron and transfer it between plasmids and chromosomes. However, most forms of *Tn5090* are now damaged and have lost the ability to move. These structures can collect numerous antibiotic resistance genes of different kinds and express them. The fact that the class 1 integron is often found on plasmids means that a bacterium can become resistance to several antibiotics with a single HGT event. Integrons are particularly important in *E. coli* for the generation of multi-drug resistance (Toleman and Walsh 2011). We now know that the Class 1 integron was responsible for some of the first cases of multidrug resistance in Gram-negative bacteria and may still be found rather often in Gram-negative pathogens (Toleman and Walsh 2011). Integrons also occur in environmental organisms and often harbour huge numbers of gene cassettes. For example, *V. cholerae* contains a super integron that includes 179 gene cassettes and forms most of *V. cholera's* second chromosome. These environmental superintegrons are the source of antibiotic gene cassettes, though, in fact, most of the gene cassettes are genes not related to antibiotic resistance (Mazel et al. 1998; Rowe-Magnus et al. 1999). In clinically relevant bacteria such as *E. coli,* the class 1 integron often includes from one to six antibiotic resistance gene

cassettes and is vitally important to multi drug resistance (Toleman and Walsh 2011).

1.13.4 IS*26*

IS*26* is one among the numerous IS that were shown to be obviously engaged in the process of resistance gene acquisition. IS*26* can accomplish this by first transposing to locations that are located on each side of a resistance gene (Terawaki et al. 1967). When an element is directly replicated either side of a resistance gene, this will make it possible for the complete structure to move in a single step, therefore mobilising everything that is located between the two copies of the gene (Mahillon and Chandler 1998). The early penicillin resistance in Gram-negative bacteria was often caused by IS*26*, which was flanked by *bla*$_{SHV}$ gene from the chromosome of *Klebsiella pneumoniae* and moved it onto plasmids that could then enter into *E. coli* strains making them penicillin resistant.

1.13.5 IS*CR* and IS*EcP-1* elements

In the early 1980s, one-ended transposition was discovered to be occurring with *Tn3* and *Tn21* for the very first time (Arthur et al. 1984; Avila et al. 1984). This transposition often relies on the misidentification of one of the transposons ends e.g., *ISEcP-1* or from the transposon having completely different ends (IS*91* and ISCR elements). The power of these elements is based on the fact that only a single copy of the insertion sequence is needed rather than two copies like a composite transposon (Mahillon and Chandler 1998).

Also, these newly emerged IS have increased the capability of mobilising genes, which is dramatically shown by the frequent presence of the IS, IS*EcP-1*, near newly-acquired resistance genes (Toleman and Walsh 2011).

IS*CR* elements are a newly-described transposon that is not related to IS*EcP-1*, and functions by rolling circle replication; a process that may mobilise extremely large sections of DNA (Toleman et al. 2006). Despite being quite different to IS*EcP-1,* it still functions in a similar way in that a single element can mobilise large sections of DNA (Toleman and Walsh 2011).

1.14 Aims and Hypothesis

The fight against infectious diseases is one of the greatest public health challenges facing our society. In particular, multidrug resistance (MDR) in *E. coli* has been listed as a critical concern by the World Health Organisation. This study hypothesised, based on the prior laboratory work around the plasmid paradox, that *E. coli* fitness is driven by the loss of plasmids since the resources needed to maintain individual resistance genes on the chromosome are considerably less than those needed to maintain a large parasitic plasmid consisting often of over 100 genes. This study hypothesised that the difference in sepsis rates could simply reflect differences in the carriage of pathogenic *E. coli* phylotypes in the community. Therefore, this study aimed to:

- Investigate the role played by chromosomal insertion of ESBL genes originating from different sources;
- Determine the common *E. coli* sequence types carried in the UK community (without antibiotic bias);
- Determine if there is any difference in *E. coli* phylogroup, MLST, and antibiotic resistance gene carriage in disparate nations; and
- By using the chromosomally located resistant strains, this study aimed to test whether strains that have lost large resistance plasmids are more virulent using an insect gut model.

Chapter 2 : Materials and Methods

2.1 Sewage samples

Raw untreated sewage was collected from wastewater treatment plants in four different countries: Saudi Arabia, Kazakhstan, Bangladesh, and the UK. This included single sites in Unaizah, Saudi Arabia and Karaganda, Kazakhstan and six sites in the UK. Sewage was also sampled in Dhaka Bangladesh, from several sewer outlet across the city. The sites in the UK included three sites towards the Eastern end of the M4 motorway (Reading, Marlow and Longreach) and three sites towards the western end of the M4 motorway serving Bristol, Newport, and Cardiff. All wastewater plants at all sites served large populations of people. Each sample consisted of c. 100 mL of crude sewage that was collected on entry to the sewage works before any treatment had commenced. Samples were mixed extensively and diluted (usually c. 2000 x) so that only five to 10 red colonies were visible on Brilliance^{TM} UTI Clarity^{TM} Agar plates (Oxoid). Samples were also centrifuged, and the pelleted bacteria were resuspended in 2 mL of freezing medium and stored in TS/72 beads (Technical Service Consultants, UK)

2.1.1 Collection dates

- Unaizah sewage work, Saudi Arabia, 06/06/2019 (one sample).
- Karaganda sewage work, Kazakhstan, 17/04/2019 (one sample).
- Newport, Ponthir sewage work, 24/06/2019 (one sample), 16/09/2019 (one sample), 24/09/2019 (one sample), 31/10/2019 (one sample).
- Bristol, Avonmouth sewage work, 19/09/2019 (one sample).
- Cardiff sewage work, 24/09/2019 (one sample).
- Longreach Marlow, Reading, all sampled 26/09/2019 (one sample each site).

- Dhaka, Bangladesh, nine sites around the city pooled, 21/10/2019 (single pooled sample) (**Figure 2.1**).

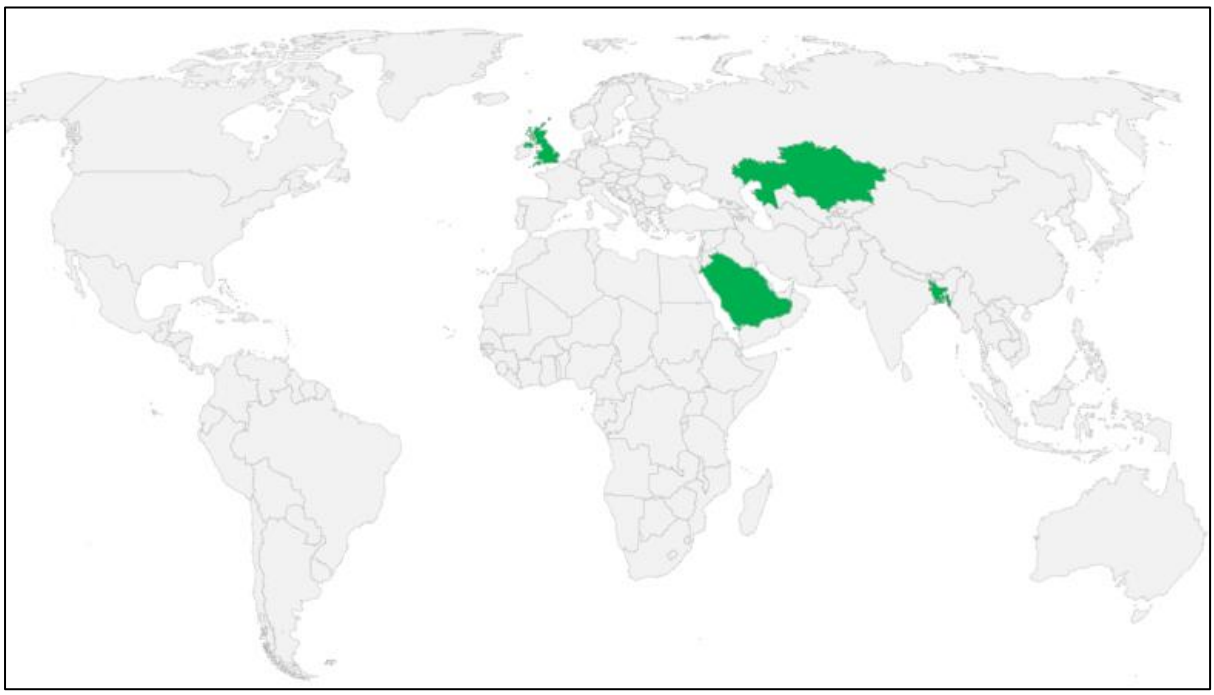


Figure 2.1 Sites of sewage plants in different countries: 1, Six sites (UK) 2, Unaizah (Saudi Arabia) 3, Karaganda (Kazakhstan) 4, Dhakka, Bangladesh

2.2 Culturing *E. coli* species

2.2.1 Sewage samples dilution and plating

Sewage samples were diluted using LB Broth, Lennox (Fisher Bioreagents™) to achieve five to 10 red colonies on a plate typically (2,000x). This was done by sequentially diluting 100 µl of well mixed sewage and plated on Brilliance™ UTI Clarity™ Agar plates (Oxoid). The plates were then incubated over night at 37°C. Brilliance plates were supplemented with cefotaxime or ciprofloxacin to give final concentrations of 10mgL-1 and 1mgL-1, respectively. Plates containing five to 10 red colonies were used and all red colonies were picked to minimise any bias in picking small or large colonies. Red colonies were further tested by the Doumith (Doumith et

al. 2012) or the updated Clermont (Clermont et al. 2013) phylotype PCR. Species identity was further confirmed by Microflex LT MALDI-TOF MS (Bruker Daltonik, GmbH, UK) with α-Cyano-4-hydroxycinnamic acid (HCCA) for confirmation.

2.2.2 Brilliance ™ UTI ™ Clarity Agar plates

Brilliance™ UTI Clarity™ Agar was used to differentiate and identify the *E. coli* isolates. Seventy-three grams of Brilliance™ UTI Clarity™ Agar (Oxoid) was resuspended in one litre of ddH_2O, Mixed and sterilised by autoclaving at 121° C for 15 minutes. The media was then cooled to 50° C and poured into sterile Petri dishes.

2.3 Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF-MS)

MALDI-TOF-MS (henceforth referred to as MALDI) was used to identify *E. coli* species as well as contaminants using a Microflex LT MALDI from Bruker, the UK. To make sure that pure single colonies were available, samples were grown freshly on blood agar plates. Then, an autoclaved wooden or plastic toothpick was used to pick out single colonies, and a small amount of bacteria was spread in a well of a MALDI plate. Using the same toothpick, but without adding more bacteria, a second spot was filled. (This is done to make sure that there is a wide range of bacterial loads to avoid wrong identification and to serve a duplicate).
One µl of MALDI Matrix (α-Cyano-4-hydroxycinnamic acid (HCCA)) (Sigma, UK) was put on the surface of each well of the MALDI plate and left to dry for 15 minutes. Then, it was processed by SACU (Specialist Antimicrobial Chemotherapy Unit at the University Heath Hospital Wales).

2.4 Polymerase chain reactions used in this project (PCR)

2.4.1 The Clermont *E. coli* phylotype PCR method

Clermont and colleagues (Clermont et al. 2000) described a multiplex PCR that amplifies several different products simultaneously in one reaction. Clermont and colleagues have described *E. coli* of consisting of four main phylogenetic groups (A, B1, B2, and D) and that virulent strains mainly belong to group B2 and to a lesser extent D. The PCR uses a combination of primers targeting three *E. coli* genes (*chuA, yjaA, and* TspE4.C2) and was originally tested with 230 known strains and showed excellent correlation with the MLEE technology. The PCR was performed following conditions: denaturation for 5 min at 94°C, 30 cycles of 30 s at 94°C, 30 s at 55°C, and 30 s at 72°C; and a final extension step of seven minutes at 72°C. The primer sequences and amplification product size are given below (**Table 2.1 & Figure 2.2**).

Table 2.1 The primers sequence and amplification product size for Clermont method (Clermont et al. 2000)

Primers	Primer sequence 5'-3'	Expected product size
chuA Forward *chuA* Reverse	GACGAACCAACGGTCAGGAT TGCCGCCAGTACCAAAGACA	279 bp
yjaA Forward *yjaA* Reverse	TGAAGTGTCAGGAGACGCTG ATGGAGAATGCGTTCCTCAAC	211 bp
TspE4.C2Forward TspE4.C2Reverse	GAGTAATGTCGGGGCATTCA CGCGCCAACAAAGTATTACG	152 bp

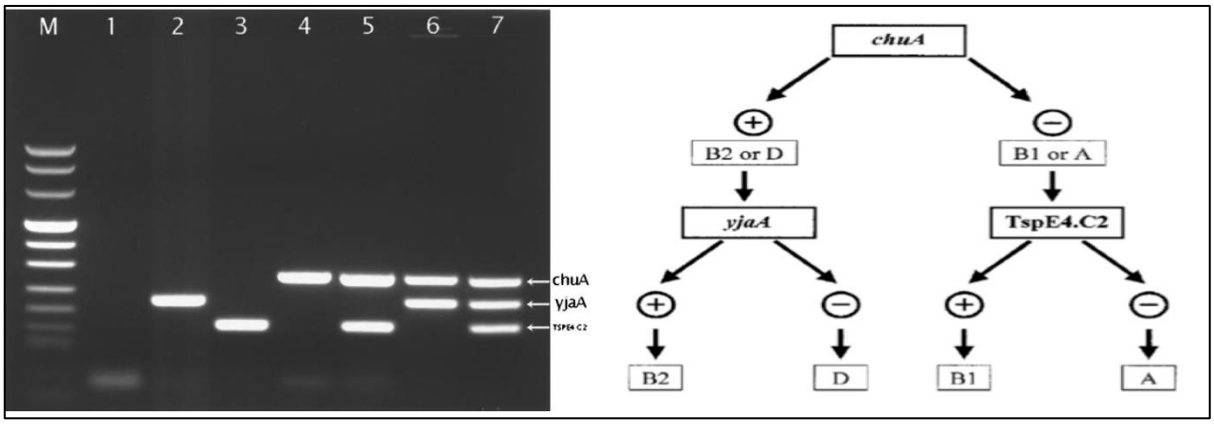


Figure 2.2: LHS Triplex PCR profiles specific for phylogenetics groups
Each combination of *chuA, yjaA* gene and TspE4.C2 amplification allowed phylogenetic group determination of individual strains. Lanes 1 and 2, group A; lane 3, group B1; lanes 4 and 5, group D; lane 6 and 7, group B2. RHS shows Dichotomous decision tree to determine the phylogenetic group of an E. coli strain by using the results of PCR amplification of the *chuA* and *yjaA* genes and DNA fragment TspE4.C2.

2.4.2 The Doumith *E. coli* phylotype PCR method (Doumith et al. 2012)

The Doumith (Doumith et al. 2012) multiplex PCR strategy is a modified version of the Clermont method that was developed because, in practice, the original technique failed to produce PCR products for some *E. coli* strains. Doumith *et al* redesigned the primers and added an *E. coli* specific PCR set as a positive control for *E. coli* (*gadA*) (Selander et al. 1986). The isolates were validated by using 185 known sequence types from multi-locus sequence typing (MLST) data.

We used this multiplex PCR using cycling conditions: (i) 4 min at 94°C; (ii) 30 cycles of 30 s at 94°C, 30 s at 65°C, and 30 s at 72°C; and (iii) a final extension step of five min at 72°C. PCR reactions included 1µL of *E. coli* culture as template, and 20 pmol

of each primer (Eurofins Genomics UK Ltd), and used 2X PCR master mix, Dreamtag^{TM} Hot start Green PCR Master Mix (Thermofisher.com) which contained dNTPs, taq polymerase and MgCl2 at appropriate concentrations and molecular grade water was added to make each reaction to 20µl total volume. The primer sequences and amplification product size are given below (**Table 2.2**).

Table 2.2 The primers sequence and amplification product size for the Doumith PCR method (Doumith et al. 2012)

	Primer sequence 5'-3'	**Expected product size**
gadA Forward *gadA* Reverse	GATGAAATGGCGTTGGCGCAAG. GGCGGAAGTCCCAGACGATATCC	373 bp
chuA Forward *chuA* Reverse	ATGATCATCGCGGCGTGCTG AAACGCGCTCGCGCCTAAT	281 bp
yjaA Forward *yjaA* Reverse	TGTTCGCGATCTTGAAAGCAAACGT ACCTGTGACAAACCGCCCTCA	216 bp
TspE4.C2 Forward TspE4.C2 Reverse	GCGGGTGAGACAGAAACGCG TTGTCGTGAGTTGCGAACCCG	152 bp

2.4.3 Updated Clermont *E. coli* phylotype PCR method (Clermont et al. 2013)

This methodology assigns *E. coli* to one of the eight phylogroups: seven (A, B1, B2, C, D, E, and F) belong to *E. colis* sensu stricto, whereas the eighth is the Escherichia cryptic clades I.

We performed PCR reactions under the following conditions: denaturation four min at 94°C, 30 cycles of five s at 94°C and 20 s at 57°C (for group E) or 59°C (for quadruplex and group C PCR), and a final extension step of five min at 72°C. The method utilises an initial quadraplex PCR that identifies most phylogroups but does not differentiate phylogroup A from C or phylogroup E. This is achieved by

secondary PCR reactions. The primers used for the allele-specific phylo-groups E and C PCRs were ArpAgpE.f and ArpAgpE.r and trpAgpC.f and trpAgpC.r, respectively. The primer sequences and amplification product size are given below (**Table 2.3** & **Figure 2.3**).

Table 2.3 The primers sequence and amplification product size for updated Clermont method (Clermont et al. 2013)

PCR reaction	Primer ID	Target	Primer sequences 5'-3'	expected Product size (bp)
Quadruplex	chuA.1b chuA.2	*chuA*	ATGGTACCGGACGAACCAAC TGCCGCCAGTACCAAAGACA	288bp
	yjaA.1b yjaA.2b	*yjaA*	CAAACGTGAAGTGTCAGGAG AATGCGTTCCTCAACCTGTG	211bp
	TspE4C2.1b TspE4C2.2bA		CACTATTCGTAAGGTCATCC GTTTATCGCTGCGGGTCGC	152bp
	AceK.f ArpA1.r	*arpA*	AACGCTATTCGCCAGCTTGC TCTCCCCATACCGTACGCTA	400bp
Group E	ArpAgpE.f ArpAgpE.r	*arpA*	GATTCCATCTTGTCAAAATATGCC GAAAAGAAAAAGAATTCCCAAGAG	301bp
Group C	trpAgpC.1 trpAgpC.2	*trpA*	AGTTTTATGCCCAGTGCGAG TCTGCGCCGGTCACGCCC	219bp
Internal control	trpBA.f trpBA.r	*trpA*	CGGCGATAAAGACATCTTCAC GCAACGCGGCCTGGCGGAAG	489bp

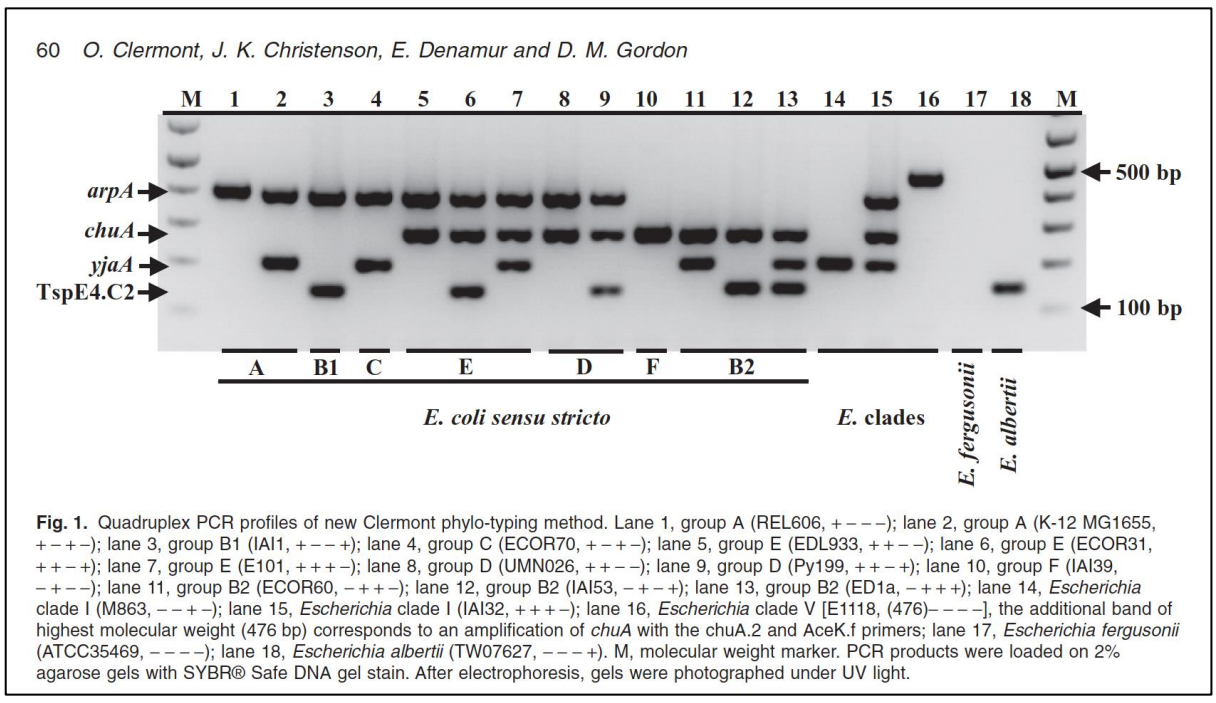


Fig. 1. Quadruplex PCR profiles of new Clermont phylo-typing method. Lane 1, group A (REL606, +−−−); lane 2, group A (K-12 MG1655, +−+−); lane 3, group B1 (IAI1, +−−+); lane 4, group C (ECOR70, +−+−); lane 5, group E (EDL933, ++−−); lane 6, group E (ECOR31, ++−+); lane 7, group E (E101, +++−); lane 8, group D (UMN026, ++−−); lane 9, group D (Py199, ++−+); lane 10, group F (IAI39, −+−−); lane 11, group B2 (ECOR60, −++−); lane 12, group B2 (IAI53, −+−+); lane 13, group B2 (ED1a, −+++); lane 14, *Escherichia* clade I (M863, −−+−); lane 15, *Escherichia* clade I (IAI32, +++−); lane 16, *Escherichia* clade V [E1118, (476)−−−−], the additional band of highest molecular weight (476 bp) corresponds to an amplification of *chuA* with the chuA.2 and AceK.f primers; lane 17, *Escherichia fergusonii* (ATCC35469, −−−−); lane 18, *Escherichia albertii* (TW07627, −−−+). M, molecular weight marker. PCR products were loaded on 2% agarose gels with SYBR® Safe DNA gel stain. After electrophoresis, gels were photographed under UV light.

Figure 2.3 Identification of specific *E. coli* phylotypes by the updated Clermont multiplex PCR methodology

2.4.4 **Woodford CTX_M extended-spectrum β-lactamase** (Woodford et al. 2006)

This PCR identifies the different CTX-M major families. Amplification conditions included: initial denaturation at 94°C for five min; 30 cycles of 94°C for 25 s, 52°C for 40 s and 72°C for 50 s; and a final elongation at 72°C for six min (Woodford et al. 2006). The primer sequences and amplification product size are given below (**Table 2.4**).

Table 2.4 The primers sequence and amplification product size for Woodford method to detect genes encoding CTX-M extended spectrum β-lactamases (Woodford et al. 2006)

Primers	Primer sequence 5'-3'	Expected product size
Group 1 forward	AAA AAT CAC TGC GCC AGT TC	415 bp
reverse	AGC TTA TTC ATC GCC ACG TT	
Group 2 forward	CGA CGC TAC CCC TGC TAT T	552 bp
reverse	CCA GCG TCA GAT TTT TCA GG	
Group 9 forward	CAA AGA GAG TGC AAC GGA TG	205 bp
reverse	ATT GGA AAG CGT TCA TCA CC	
Group 8 forward	TCG CGT TAA GCG GAT GAT GC	666 bp
reverse	AAC CCA CGA TGT GGG TAG C	
Group 5 forward	GCA CGA TGA CAT TCG GG	327 bp
reverse	AAC CCA CGA TGT GGG TAG C	

2.5 Gel electrophoresis

2.5.1 PCR agarose gel preparation

Two-hundred mL of 1X Tris/Boris/EDTA (TBE) was poured into a Duran laboratory bottle and 3g of Hi-Res standard agarose (1.5%) were added and microwaved for two mins to ensure agarose crystals are dissolved. Then, 20µl of 10mg.$^{L-1}$ Ethidium bromide was added to the molten agarose, and the agarose poured into a gel mould. The gel was subsequentially left on the bench for ~20 minutes to solidify before use.

Gels were run at 250V for 45 minutes in a gel electrophoresis tank with TBE (1X) as the electrolyte. Once completed, the gel was visualised using a UV transilluminator.

2.6 Antibiotic susceptibility testing

Isolates were assayed by discs diffusion (Oxoid) using discs containing 14 different antibiotics and antibiotic/inhibitor combinations. These included: ciprofloxacin (CIP, 1µg) sulphonamides: sulphamethoxazole/trimethoprim (SXT, 1,25/ 23,75µg) aminoglycosides: gentamycin (CN, 30µg) and β-lactams: amoxycillin (AML, 20µg) cefotaxime (CTX, 30µg) ceftazidime (CAZ, 30µg) tazobactam/piperacillin (TZP, 10/75µg) and amoxycillin/clavulanic acid (AMC, 20/10µg), chloramphenicol (CL, 30µg), mecillinam (MIL, 10µg), fosfomycin (FOT, 200µg), nitrofurantoin (F, 100µg), trimethoprim (TM, 5µg), amikacin (AK, 30µg) (**Table 2.5**). Turbidity was adjusted in 0.85% sterile normal saline solution to make 0.5 McFarland's standard and a lawn of bacteria was then seeded on MH agar (ThermoFisher Scientific Inc) with a sterile cotton swab. Discs impregnated with antibiotics were placed on the plate and the plates were incubated for 24h at 37°C. Breakpoint were measured using a printed template of the zones of inhibition and recorded as sensitive, resistant, or intermediate following CLSI guidelines. Control strains were *E. coli* strains. For example: CIP sensitivity >14mm and intermediate <16, AML sensitivity > mm and intermediate < mm, TZP sensitivity >14 and intermediated <17, AMC sensitivity >15 and intermediate < 17, CTX sensitivity >14 and intermediate <16, FOS <24 and intermediate <24.

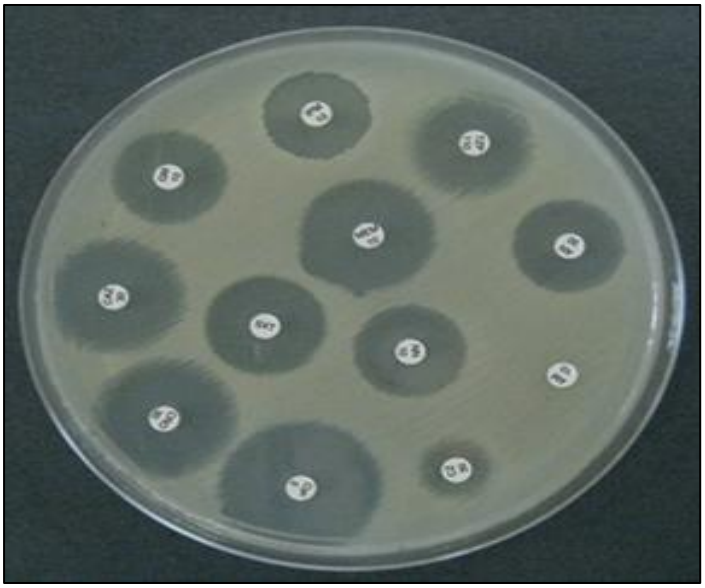


Figure 2.4 Example image of disc diffusion assay
A lawn of *E. coli* is streaked, and discs impregnated with antibiotics are placed atop the agar before incubating at 37°C for 24 hours.

The zone sizes were determined by overlaying the plate on a predefined template that gave zones of inhibition for resistant, sensitive, and intermediate diameters. Results were determined following CLSI guidelines (**Table 2.5**).

Table 2.5. Zone diameter of antibiotics used in this study

Antibiotic	Disk contents µg	Zone Diameter (mm)		
		Resistant	Intermediate	sensitive
Cefotaxime (CTX)	30	≤ 14	14-16	≥ 17
Ceftazidime (CTZ)	30	≤11	12-19	≥20
Ampicillin (AMP)	20	≤13	13-16	≥17
Amoxicillin/Clavulanic (AMC)	20/10	≤ 15	15-17	≥18
Piperacillin/tazobactam (TZP)	10/75	≤14	14-17	≥18
Ciprofloxacin (CIP)	1	≤14	15-16	≥16
Fosfomycin (FOS)	200	≤24	-	≥25
Amikacin (AK)	30	≤18	-	≥18
Gentamycin	30	≤17	-	>17
Nitrofurantoin	100	≤11	-	>11
Mecillinam	10	≤15	-	>15
Trimethoprim	5	≤15	-	>15

2.7 S1 Pulsed Field Gel Electrophoresis (S1-PFGE).

S1-PFGE and DNA restriction was conducted according to Standard Operation Procedure for PulseNet PFGE of *Escherichia coli* o157:H7, *Escherichia coli* non-O157 (STEC), Salmonella serotypes, *Shigella sonnei*, and *Shigella flexneri* gels (Available at (CDC 2024). Gels were subsequentially processed and probed using S1 in gel hybridisation protocol (Toleman 2018).

2.7.1 Preparing agar plugs

Bacterial isolates were plated on as a lawn on Columbian blood agar (CBA) plates (supplied by Oxoid) at 37°C overnight. A 10 mL loop was used to scrape about two

inches of confluent growth from the CBA plate and resuspended in two mL of cell suspension buffer (CSB) and adjusted to make a final concentration of OD_{600} 0.8-1.0. Four-hundred µl of CSB was subsequently mixed with 20 µl of Proteinase K (20 mg/mL) and 400 µl of melted 1% SeaKem Gold agarose (SKG), (Lonza, Switzerland). It was mixed gently and rapidly dispensed into clean plug mold to prevent solidification. Solidified plugs were then transferred into polypropylene screwcap after adding five mL of cell lysis buffer and 25 µl of Proteinase K (20 mg/mL) and incubated at 55°C at 150 rpm in a shaking incubator for two hours. The plugs were then washed twice with 10-15 mL pre-heated (54-55°C) double distilled water, and with 10-15 mL of pre-heated (54-55°C) TE buffer for four times. Each wash was kept in a shaking incubator at (54–55°C) for 20 minutes. Plugs were then stored at 4°C with TE buffer. For S1-PFGE compositions please see appendix 1.

2.7.2 Restriction digestion of DNA in casted plugs

Each plug was cut 2.0-2.5 mm in size and washed in 300 µl of 1/10 of TE buffer and incubated in a RT for 5 minutes. Then, in 200 µl of 1X S1 buffer for 5 minutes at RT. The S1 enzyme was diluted 1/20,000 by adding 0.5 µl of (50U) of S1 (stock enzyme at a concentration of 100U/µl) to 10 ml of 1X S1 buffer. Digestion was done by adding 200 µl of diluted S1 enzyme (Invitrogen, the UK) to each plug and incubating it at 4°C overnight.

2.7.3 Casting an agarose gel

PFGE system suppliers' gel-forming equipment was used to make 0.9% agarose gels by adding 1 gm of PFGE grade agarose to 110 mL of 0.5X TBE and microwaving it for two minutes. The gel was left to solidify for 20 minutes before loading the plugs. The gel was then put on the PFGE equipment under the gel

running buffer of 2L of 0.5X TBE. On a CHEF II PFGE machine (Bio-Rad), the gel running parameters included: initial switch time: 5s, final switch time: 45s, voltage: 6 V, run time: 22 h, included angle: 120°. A Lambda concatemer PFG ladder (NEB, UK) was used as a molecular size marker and 50 µL of a 10 mg/mL solution of ethidium bromide was added to the gel running buffer. Plasmids were visualised by placing on a UV transilluminator, and a UVP geldoc II imaging system was used to take pictures of gels (UVP Cambridge, the UK).

2.7.4 In gel hybridisation

Gels were initially dried in a 50°C drying cabinet overnight by putting them between two sheets of blotting paper (Toleman 2018). Then, rehydrated into 200 mL of deionised DNase-free water in a flat-bottomed Pyrex glass bowl for 5 minutes. Rehydrated gels were subsequently denatured using, 200 mL of denaturing solution for 45 mins at room temperature and lastly neutralised by removing the denaturing solution and adding 200 mL of neutralising solution for 45 minutes. The gels were then removed and placed in a hybridisation tube and incubated with 20 mL of pre-hybridisation solution at 65°C for at least 72h.

Prehybridisation solution (20 mL) was made from stock solutions of: 5% PVP (0.4ml), 5% ficoll (0.4ml), 10mg/ml herring testes DNA (0.3ml), 20X SSC (6mL) and 1mL of full cream UHT milk. This was adjusted with DNase free water to 20 mL.

Probes (CTX-M-15) were prepared by amplifying a resistance gene *bla*$_{\text{CTX-M-15}}$ from a *Klebsiella pneumonia.* The PCR product was subsequently extracted from the gel and purified (QIAquick Gel Extraction Kit, Qiagen, Germany). Identity was confirmed by sequencing and then the PCR product was labelled using a stratagene random

priming labelling kit with ^{32}P dCTP (Stragene, Amsterdam, Netherlands) using the manufacturer's instructions.

Once the probe had been labelled, unlabelled ^{32}P dCTP and nucleotides were removed using a Sephadex G50 gravity flow gel filtration column (illustra™ Nick™ Columns Sephadex G-50 DNA grade, GE Healthcare Life Sciences, little Chalfont, Buckinghamshire, the UK). This was done by adding 60 µL of the labelled probe to the top of the gel filtration column and waiting for it to enter the gel. Then 320 µL of 0.1 M Tris pH 7.5 is added to the top of the column. Gravity moves the filtrate through the column, where it is collected in a 1.5 mL Eppendorf tube and disposed. Then, a new Eppendorf tube is put under the column, and 430 µL of 0.1 M Tris pH 7.5 is used to elute the labelled DNA. This leaves unincorporated nucleotides on the column. Then labelled probe is then boiled for six minutes in an Eppendorf tube with a screw cap and added to the prehybridised gel and left to hybridise at 65 °C overnight. After hybridisation, the probe is discarded, and the gel is washed for one hour at 65°C with 100 mL of 2X SSC 0.1% SDS and then for one hour at 65°C with 0.1 SSC 0.1% SDS (100 mL). Gels that have been washed are finally taken out of the hybridisation tubes, washed for a few minutes under a warm tap, and then dried with blotting paper. Then, they are wrapped in cling film and put against a sheet of film (Lumi-Film Chemiluminescent Detection film, Roche, Mannheim, Germany) in a film cassette overnight at a freezer -80. The next day, standard film development and fixer solutions are used to develop the film.

2.8 Illumina sequencing

2.8.1 DNA extraction

Bacterial isolates were plated on CBA at 37°C overnight. A single colony was then resuspended in 1.8 mL of LB broth and grown overnight in a shaking incubator (100 rpm at 37°C). After overnight incubation the same was centrifuged at 10000 rpm for 10 minutes, the supernatant removed and loaded onto an automated QIAcube machine (QIAGEN), run with QIAamp DNA mini kit (QIAGEN, Germany).

2.8.2 DNA quantification

The Qubit™ 4.0 fluorometer (ThermoFisher Scientific, the UK) and the dsDNA 1X Kit were used to quantify the amount of genomic DNA. When compared to the dsDNA narrow range kit, the 1X kit is a mix of buffer and dye that has already been made to reduce lab waste and speed up use. At first, two control tubes with 190µL of 1X dye and 10µL each of standard 1 and standard 2 were set up. Standard 1 is a negative control and is 0 ng/µl and standard 2 is appositive control of 10ng/µl. The addition of these two standards creates a calibration curve so samples of unknown DNA concentration can be determined from the fluorescence they produce. The samples were measured by adding 2µL of sample DNA extract to 198µL of 1X dye into a 0.5mL Eppendorf tube, vertexing for 2-3 seconds, and then inputting the mixture into the Qubit™ 4.0 fluorometer (ThermoFisher, the UK).

2.9 Oxford nanopore technology (MinION)

2.9.1 Rapid barcoding kit 96 (SQK-RBK110.96)

The rapid barcoding kit has these contents: rapid barcode plate (RP96), AMPure XP beads (AXP), sequencing buffer II (SBII), rapid adaptor F (RAP-F), elution buffer (EB), loading beads (LBII), loading solution (LS), flush tether (FLT), flush buffer (FB) with different cap colour for each bottle. All components are first thawed at RT. The DNA was prepared in PCR tubes using nuclease free water by adjusting the volume to nine µl of 50ng genomic DNA and mixed by pipetting up and down and spun down using a microfuge. Using a multichannel pipette 1µl of Rapid Barcodes are transferred to the PCR tubes to get a total of 10µl. The solution is then mixed by pipetting up and down and spun down using a microfuge. The PCR tubes were incubated in a thermal cycler at 30°C for two mins and then at 80°C for two mins and cooled on ice for two mins. This enables the barcodes to be attached to the DNA by a transosome complex. All barcoded samples were then pooled into 1.5 mL Eppendorf. The mixed DNAs are then purified using magnetic beads (SPRI beads) with 80% alcohol. This consisted of two washes of 1.5 mL. After the second ethanol wash, the alcohol was removed, and the tube was set aside for up to five minutes to dry in the air and any residual ethanol was removed using a pipette. The dried pellet was mixed with between 15 µL of elution buffer. This mixture was then shaken for two mins at 1800rpm, left at room temperature for two minutes, and placed on the magnet until clear (around five mins). The 15µl of elute was retained and transferred to a clean 1.5 mL Eppendorf DNA tube. The pellet bead was discarded and 11 µl of the sample was transferred to 1.5 mL Eppendorf DNA tube. To this tube 1 µl of

Rapid Adaptor F (RAP F) was added and mixed by flicking the tube and incubated at RT for five mins.

2.9.2 Loading the flow cell preparation

The lid and sliding port of the MinION were opened so that the priming port could be seen. Using a pipette, a small amount of the buffer was taken out to remove any bubbles and to avoid damaging the pores. This made sure that the priming port got a steady flow of buffer. The flow cell priming mix was prepared by adding 30 µl of FLT buffer to 1.17 mL of FB. Then the port was filled with 800µL of the priming mix (FB+FLT). This was then left for five minutes, during which time the DNA was prepared. The DNA library was prepared in a new tube by adding 37.5 µl of SBII, 25.5 µl of LBII to 12 µl of the DNA library. After five minutes, the SpotON sample port cover was removed, and 200 µL of the priming mix (FB+FLT) was loaded into the flow cell through the priming port. The final DNA Library was then mixed and added drop by drop through the SpotON sample port. Then, the MinION's ports and lid were closed. And using the MinKNOW GUI, by following the instructions on the screen and choosing the SQK-RBK110.96 kit. The minion was run for 72 hours until sequencing was complete.

2.9.3 Basecalling using Guppy

Nanopore fast5 files were transferred to the Cardiff Hawk supercomputer and basecalled using Guppy (v5.0.11). The converted FASTQ files were assembled using Flye software (v2.8.1), and annotated using Prokka.

2.10 Whole genome Sequencing (WGS)

Following the instructions of the manufacturer, Nextera XT v2 (Illumina, USA) genomic libraries were created utilising a bead-based normalisation. Utilising the v3 chemistry, paired-end WGS was carried out on an Illumina MiSeq to produce fragment lengths up to 300 bp (600 cycles).

2.10.1 DNA preparation

The amount of DNA was set to 0.2ng/mL. This was done by adding the Qubit value (DNA concentration) and the sample ID to make a 0.2ng/L concentration plate by adding the volume of DNA in µL to the volume of mgH2O in µL. The full plate was then mixed using a multichannel, and bubbles were taken out by spinning it down.

2.10.2 Tagmentation, PCR and clean up

A 96-well plate was made with 5µL of TD buffer (Tagmentation DNA Buffer), 2.5µL of ATM (Amplification Tagmentation Mix), and 2.5µL of 0.2ng/µL DNA that had just been mixed and spun down. Then, this was mixed, spun down, sealed with a PCR seal, and put in a thermocycler that had already been heated to 55°C for five mins. Then, 2.5 µL of a buffer called "Neutralise Tegument Buffer" (NT) was added, mixed, and left for five mins at room temperature to stop any more DNA from being tagged.

Also, 7.5µL of NPM (Nextera PCR Master Mix) was added to the 96 well plate. Each isolate was given a unique indice combination from one of the four sets of indices (A, B, C, & D). This allows for fragments of DNA pertaining to a specific isolate to be barcoded and thus enables demultiplexing post sequences. Mixing, sealing, spinning, and running the PCR conditions are as follows (**Table 2.6).**

Table 2.6. PCR condition of primer used in tagmentation

Temperature °C	Time (seconds)	Repetitions
72	180	1
95	30	1
95	10	12
55	30	
72	30	
72	300	1
10	∞	1

2.10.3 PCR clean up and library normalisation

Each well of a 0.8mL MIDI plate labelled "Clean Amplified Plate" was filled with 11.5µL of SPRI (Solid Phase Reversible Immobilisation) beads.

SPRI beads were pre-warmed to 30°C because the yield from the process would be much lower if they were not. For the beads to be spread out evenly in the liquid, the liquid must be vortexed for at least 30 seconds. The plates were then sealed and spun (pulsed for 10 seconds at 1,000 rpm), and then they were shaken for two mins at 1,800 rpm on a VWR DMS-2500 high-speed microplate shaker. After five mins at RT, the tubes were put on a neodymium magnetic stand for two mins and the supernatant was discarded. A wash with 80% ethanol (100µl per sample) was done twice in the well. After the plate was rinsed, it was left to dry for 15 mins.

Eluted into 25 µL of RSB (Resuspension Buffer), shaken for two mins at 1,800 rpm, incubated at room temperature for two mins, and put on a magnetic plate for two mins. The supernatant was eluted into a new 96-well plate (CAN plate).

2.10.4 Library normalisation

A new 96-well plate was made with 22.5µL of LNA1 and LNB1 in each well. This was done by mixing 2,200µL of LNA1 and 400µL of LNB1 for 96 samples at a ratio of 5.5:1.

In each well, 10µL of sample from the previous plate was put in. The LNP plate was shaken for 30 mins at 1,800rpm.

The LNP plate was put in the magnetic stand for two mins, the supernatant was thrown away, and then 22.5µL of LNW1 was added to each well of the LNP plate, which was then sealed, spun down, and shaken at 1,800rpm for five mins. The plate was then put back in the magnetic stand, the supernatant was thrown away, and the LNW1 wash was done again. Then, 0.1M NaOH was made fresh. (10M NaOH solution: 4,950 µL of MgH_2O and 50 µL of 10M NaOH). Each well got 15µL of NaOH and was shaken for five mins at 1,800rpm (this allow for the DNA to be eluted from the magnetic bead). After two mins, 10 µL of the supernatant was put on a magnetic rack and put into a 96-well plate with 10 µL of LNS1 (thus making the SGP plate). Then it spun at 1,000rpm for one minute.

2.10.5 Loading MiSeq

5µl from each well of the SGP plate was combined to make the pooled amplicon library (PAL). 24 µl of PAL was then diluted with the addition of 576µl HT1 buffer to make the diluted amplicon library (DAL). Based on prior measurement of the DNA concentration of the CAN plate, a specific volume of DAL was chosen for discarding. This is to ensure suitable cluster density on the run. All samples were put into one 1.5mL eppendorf, and 24µL of it was moved to another 1.5mL eppendorf that had 576µL of HT1 in it. 100 µL of it was thrown away, and 10 µl of 20pmol of PhiX was

added, then filled up to 600µL with HT1. It was turned upside down for 150 seconds at 96°C, and then put in an ice bath. The sample was then put into a MiSeq reagent cartridge that had been taken out of the freezer and tapped down. The reagent cartridge was then put on the MiSeq, and the instructions on the screen were followed. Sequences from the MiSeq were processed in house using the Hawk supercomputer and assembled using Spades. Sequences were annotated using Prokka and antibiotic resistance genes were identified using Amrfinder.

2.11 Measurement of growth rates of plasmid and chromosomally encoded *bla*$_{\text{CTX-M-15}}$ gene in *E. coli*

Seventeen *E. coli* isolates were divided into three groups for comparisons (**Table 2.7**). A scrape from the top of frozen beads -80°C of each strain was taken using a sterile loop and used to inoculate Luria-Bertani Agar (LB) plates which contained 10 mg/L ceftazidime. After ~18 hours of incubation at 37°C, one colony was picked and inoculated into a 2 mL tube containing 2 mL of LB broth. This tube was incubated at 37°C on a on a rotary shaker at 160 rpm for two hours. An aliquot (1µl) of the bacteria cells was inoculated into (999 µl) LB broth and serial dilution made by adding (100µl) of bacteria cells to (900µl) LB broth to get a final concentration of 10^{-5} cells/mL. Concentrations were confirmed by dilution and overnight growth on LB plates. A (10µl) aliquot of the diluted acclimation culture (10^5 cells/ml) was added to the wells of microtiter plate, each containing (190µl) of LB broth. These cultures were incubated at 37°C in a power wave 600 (FLUOstar Omega) plate reader, which automatically records the optical density (OD) every thirty minutes, with the plate being shaking between readings. The last five wells were blank by only adding (200µl) LB broth. Every strain was repeated five times to get the median of each

strain. Maximum growth rate was measured by regressing OD values versus time (FLUOstar Omega microplate Reader software).

2.11.1 Strains and cultures

Seventeen *E. coli* ST131 that contained the *bla*$_{CTX\_M\_15}$ gene were used in this experiment (**Table 2.7**). These were of three groups, dependant on the genetic location of the *bla*$_{CTX\text{-}M\text{-}15}$ gene which was determined by both in gel hybridisation and in-silico sequencing in Chapter 3. These groups were plasmid or, chromosome only (groups 1&2), or isolates which carried the gene on both chromosome and plasmid. *E. coli* strains were cultured on Columbia Blood Agar plates (CBA) (supplied by Oxoid) for 24 h at 37°C. For infections, bacteria were subcultured into 6 mL Muller-Hinton Broth (MH) (supplied by Oxoid) and grown in a shaking incubator at 160 rpm and 37°C for two hours. Two millilitre of each culture was centrifuged for six minutes at room temperature at 6000×, and this procedure was repeted twice. Supernatant were discarded, and the pellet was re-suspended in 1 mL of phosphate buffered saline (PBS). The bacteria were adjusted to OD_{600nm} of 1 in PBS (pH 7.2). For infections, equivalent to 1×10^6 colony forming unit per mL (CFU/mL). Inoculation was always confirmed by viable counts, and infections at lower doses were adjusted accordingly.

2.11.2 *Galleria mellonella* survival

For infections, bacteria were subcultured into 6 mL Muller-Hinton Broth (MH) (supplied by Oxoid) and grown in a shaking incubator at 160 rpm and 37°C for two hours. Two millilitre of each culture was centrifuged for six minutes at room temperature at 6000×, and this procedure was repeted twice. Supernatant were discarded, and the pellet was re-suspended in 1 mL of phosphate buffered saline

(PBS). The bacteria were adjusted to OD_{600nm} of 1 in PBS (pH 7.2). For infections, equivalent to 1×10^6 colony forming unit per mL (CFU/mL). Inoculation was always confirmed by viable counts, and infections at lower doses were adjusted accordingly. The larval were purchased from (Live Foods Ltd) and maintained on wood chips in the dark at room temperature. Healthy larvae were defined as those possessing a uniform cream colour that lack discoloration (melanisation), high motility with the ability to right themselves when turned over, approximately 250 mg (+/- 10%) in weight, and 2–3 cm in length (**Figure 2.5**)(Jemel et al. 2020). The infection of larvae was carried out as previously described (Senior et al. 2011), using a micro-injection technique whereby 10 µl of *E. coli* strain was injected into the haemocoel via the right foreleg, using a Hamilton syringe (**Figure 2.5**)(Jemel et al. 2020). Larvae were then incubated in the dark at 37°C and survival and macroscopic appearance were recorded at 24 h post-infection. Survival of infected larvae (n=3X 10 per group) was recorded every 24 hours for 72h. Larvae were considered dead when they turned dark brown, black in colour or failed to respond to touch. For all the experiments: positive groups were injected with an extensively drug resistant *K. pneumonia* sequence type 16 (ST16) and negative groups were injected with sterile phosphate buffered saline (PBS). The last group was a mock injection (by stabbing the larvae using Hamilton syringe) to ensure that the injection procedure was not causing death.

Table 2.7 Identification and numbers of selected *E. coli* ST131 strains in three groups

Plamidic and chromosomal location	Plasmidic location	Chromosomal location
I-705	2224	2221
1709	0766	2223
2216	2203	231
212	2228	2229
1538	2197	1540
795	1684	

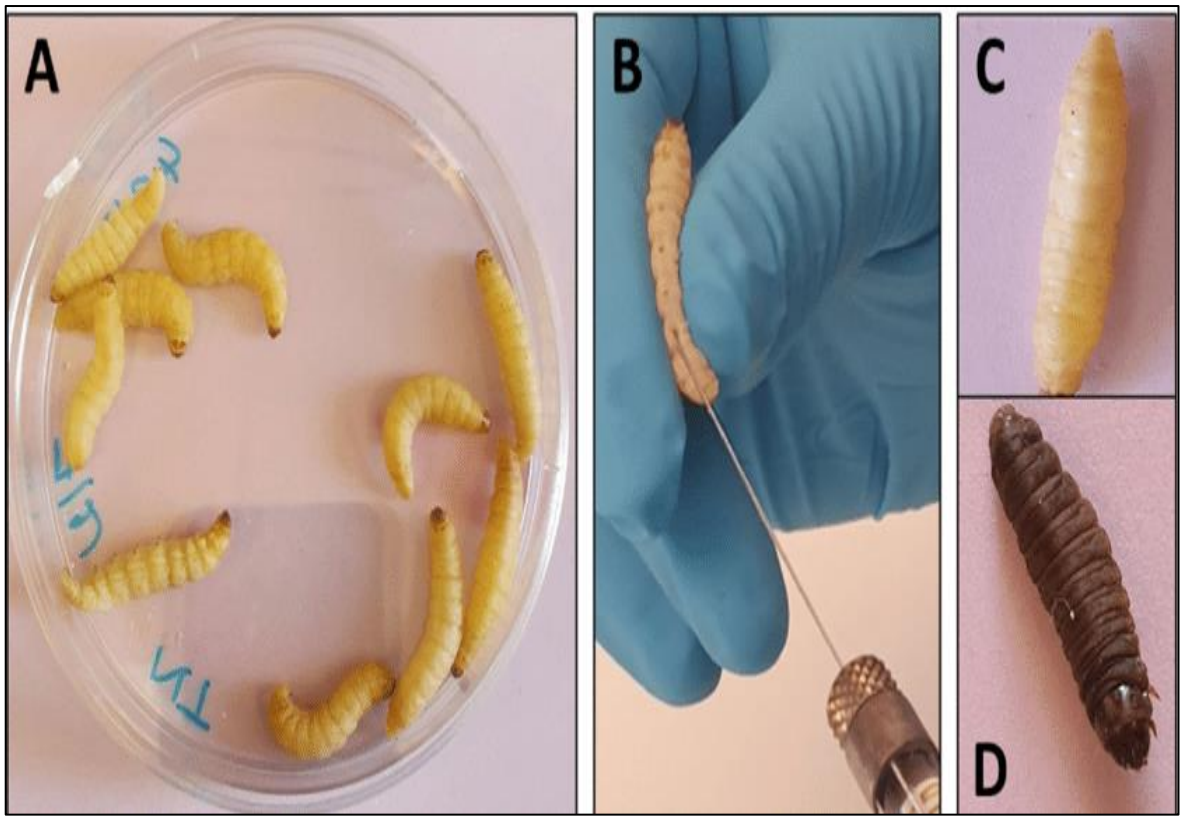


Figure 2.5 Use of Galleria mellonella larvae for evaluation of the virulence of *E. coli* ST131

(A) Larvae are grouped in ten larvae in a petri dish. (B) using the Hamilton syringe to inject the larvae into the ventral face of the last proleg. (C) living larvae. (D) Dead and melanized larvae (Jemel et al. 2020)

2.11.3 Statistical analysis

Statistical analysis of the *Galleria* experiments was performed using GraphPad Prism software (Version 9.4.1, San Diego, CA, USA). The Gompertz model may be regarded as the best model to represent the growth data in practically all circumstances (Zwietering et al. 1990). The nonlinear regression (curve fit) was performed by using the modified Gompertz equation to detect the growth rates.

$$y = A\exp\left\{-exp\left[\frac{\mu m \cdot e}{A}(\lambda - t) + 1\right]\right\}$$

It is frequently helpful to show the logarithm of the relative population size since bacteria multiply exponentially [y = ln (N/N_0)] against time. Where the three phases of growth curve can be described: which they are the maximum growth rate, μm, the lag time λ, and the asymptote [A = ln (N_∞/N_0)], which is the maximal value reached (Zwietering et al. 1990).

GraphPad Prisms nonlinear regression routine was used to fit the equation. With a 95% confidence interval, the algorithm was chosen from the set of parameters with the lowest sum of squares for the residuals. The growth kinetics, including the exponential growth rate, were then calculated using the Gompertz parameters (A, B, C, and M). Two-way ANOVA followed by Tukey's multiple comparison tests to compare the growth rates of the three groups.

To evaluate the lethality of different strains survival curves were made using the Kaplan-Meir method. The log-rank (Mantel-Cox) test was performed to analyse the survival curves on *E. coli* ST131 strains. *P* values of <.05 were considered statistically significant.

Chapter 3 : Chromosomal Integration of *bla*$_{\text{CTX-M-15}}$ Genes within a Contemporary UK National Study Collection of ESBL *Escherichia Coli*

3.1 Introduction

3.1.1 Origin of *bla*$_{CTX-M-15}$ in India

The extended-spectrum β-lactamase (ESBL) CTX-M-15 is arguably the most successful and important resistance mechanism that has emerged in *E. coli* since the start of the antibiotic era. It belongs to the CTX-Munich family of serine Ambler class A β-lactamases. The ESBLs are enzymes that can hydrolyse penicillins, expanded-spectrum cephalosporins (such as cefotaxime, ceftriaxone, ceftazidime, or cefepime) and monobactams (aztreonam), but are unable to hydrolyse carbapenems and are neutralised by β-lactamase inhibitors such as clavulanic acid and tazobactam (Cantón et al. 2012). The CTX-M family comprises over 250 variants of which CTX-M-15 is the most common (Naas 2017). It was initially discovered in cefotaxime and quinolone resistant *Klebsiella pneumonia*, *E. coli* and *Enterobacter aerogenes* isolates in New Delhi in India in 1999 (Karim et al. 2001) and quickly became the dominant resistance mechanism, being found in *E. coli* and *Klebsiella pneumoniae* clinical isolates in numerous medical institutions throughout India in 2000 (Mathai et al. 2002). The *bla*$_{CTX-M-15}$ gene was often found in isolates with chromosomal mutations giving rise to quinolone resistance, and with co-carriage of the *bla*$_{OXA-1}$ b-lactamase and aminoglycoside resistance mechanisms in these initial *E. coli* isolates (Toleman MA, Beidenbach D 2002; Walsh et al. 2007). The combination of these resistance mechanisms meant that *E. coli* were routinely resistant to most antibiotics and antibiotic/inhibitor combinations leaving the carbapenems as the only viable treatment option. Since 2000 *E. coli* isolates possessing the CTX-M-15 mechanism have spread throughout the world and become the dominant ESBL within *E. coli* in many nations (Cantón et al. 2012). In

2008, several groups identified a common *E. coli* sequence type (ST), ST131 spreading throughout Asia, Europe and Northern America that was routinely quinolone resistant and often carried the *bla*$_{\text{CTX-M-15}}$ gene, and to a lesser extent co-carriage of *bla*$_{\text{OXA-1.}}$ (Clermont et al. 2008; Coque et al. 2008; Nicolas-Chanoine et al. 2008). Recently, the initial Indian *E. coli* isolates collected in 2000 were sequenced revealing that *E. coli* ST131 producing CTX-M-15 was widely dispersed in Indian hospitals since 2000 (Paul et al. 2020). Early emergence and dissemination of *Escherichia coli* ST131 H30RX from India (Babenko et al. 2019). CTX-M-15 is also dominant in many other species of Gram-negative bacteria notably *Klebsiella pneumoniae*, *Proteus mirabilis* and other Enterobacterales.

3.1.2 *bla*$_{\text{CTX-M-15}}$ in the UK

Before 2000 most ESBL producers referred to the Public Health Agency (PHA) from England, Wales and Northern Island were hospital *Klebsiella* spp. producing TEM and SHV variants. For *E. coli,* quinolone resistance was below 5% and resistance to 3rd generation cephalosporins was rarely found (c. 1%). This situation changed quite quickly after 2001 with *E. coli* bacteraemia isolates resistant to both quinolones and 3rd generation cephalosporins appearing for the first time in 2001 and then expanding to above 10% of all *E. coli* isolated from bacteraemia's by 2007 (**Figure 3.1**). A large UK study (Day et al. 2019) has recently reported that 2,157 (11%) of 20,243 human faeces samples collected in five regions of the UK (London, Wales, Scotland, East Anglia and the North-West) contained ESBL-*E. coli*, including 678 (17%) of 3,995 in London, indicating carriage of ESBL-*E. coli* is a common occurrence in the UK. The same study also recorded ESBL-*E. coli* were frequent in sewage and retail chicken meat samples 65% (104/159), though rare in other foods. *E. coli* ST131 was the dominant sequence type found in 64% (188/293) of human

bacteraemia isolates, 36% (128/360) of faecal isolates and 22% (14/65) sewage isolates with *bla*$_{CTX-M-15}$ being the dominant ESBL gene (77% 319/416). Interestingly, *E. coli* ST602, 23 and 117 with the *bla*$_{CTX-M-1}$ ESBL gene dominated among food and veterinary isolates indicating almost no links between food and human carriage or disease due to ESBL-*E. coli.*

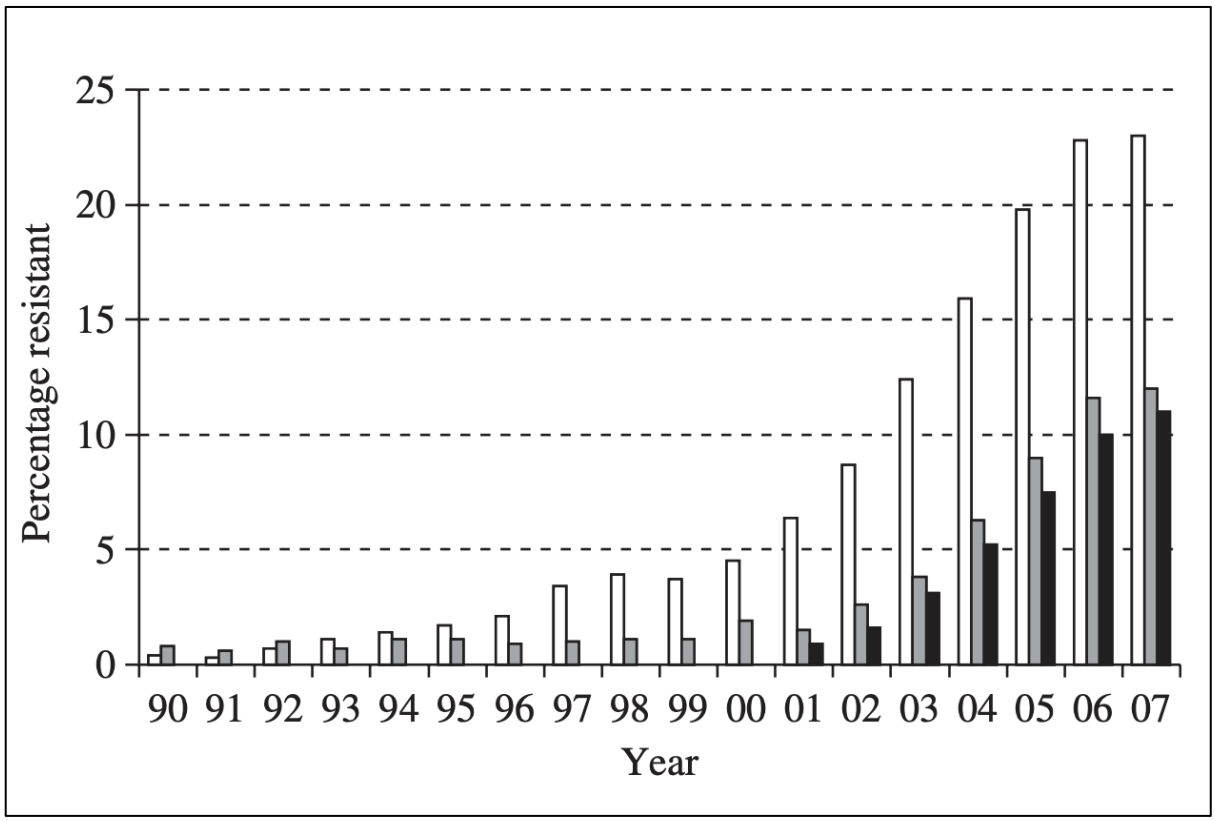


Figure 3.1 Rise in the proportions of antibiotic-resistant *E. coli* from bacteraemias

Figures for England, Wales, and Northern Ireland of isolates resistant to fluoroquinolones (white), oxyimino-cephalosporins (grey) and both (black). Based on laboratories' reports to the HPA. Taken from "Has the era of untreatable infections arrived?" David M. Livermore Journal of antimicrobial chemotherapy (2009) 64, Suppl. 1, i29-i36 doi:10.1093/jac/dkp255

3.1.3 *bla*$_{CTX-M-15}$ genetic support and the plasmid paradox

Since its first description, the *bla*$_{CTX-M-15}$ gene has been closely associated with the insertion sequence IS*Ec*P1, which was found 48bp upstream of *bla*$_{CTX-M-15}$ within large plasmids in the initial isolates and providing promoter sequences that drive

bla$_{CTX-M-15}$ expression. The *bla*$_{CTX-M-15}$ gene likely originates from the chromosome of *Kluvera ascorbata* which contains a gene identical to *bla*$_{CTX-M-3}$ and differs from *bla*$_{CTX-M-15}$ by only a single nucleotide. The IS*Ec*P1 element was likely responsible for the initial capture of *bla*$_{CTX-M-15}$ from the *Kluvera ascorbata* chromosome and its incorporation into resident conjugative plasmids allowing its escape from the chromosome and then cells of *K. ascorbata.* These capture and transfer events between *Kluvera* and *E. coli* have been replicated in lab-based experiments (Poirel et al. 2005).

Whilst transposable elements such as IS*Ec*P1 are important genetic vehicles that transport genes within bacterial cells such as between the chromosome and resident plasmids they are unable to transfer genes between bacterial cells. There are several known mechanisms of transfer of genetic information between bacterial cells including transformation, transduction, and conjugation. Transduction is of limited utility since bacteriophage tend to be very specific and generally only able to transfer DNA between similar strains within a bacterial species. Similarly, transformation is often limited to the uptake of DNA from closely related bacterial strains. Thus, horizontal gene transfer is commonly driven by conjugative plasmids and Integrative and conjugative elements (ICE) which transfer cytoplasm to cytoplasm and chromosome to chromosome, respectively (Toleman and Walsh 2011). Promiscuous conjugative plasmids that can transfer antibiotic genes between bacterial species are especially important. Whilst ICE are an important means of antibiotic gene transfer in some bacteria such as SXT in *Salmonella* and *Vibrio cholera* (Toleman and Walsh 2011) the transfer of *bla*$_{CTX-M-15}$ has mostly been attributed to conjugative plasmids in *E. coli*, which are typically of the *inc*F II type.

Conjugative plasmids are essentially bacterial parasites that impose a fitness burden on their host. This is due to the physiological costs of replicating their DNA and expressing their genes, and results in their host bacterium becoming less able to compete with plasmid-less bacteria for limited resources. The existence of conjugative plasmids has been seen by some researchers as a paradox since non-beneficial plasmids should be lost to purifying selection and beneficial genes carried on plasmids should be incorporated into the bacterial chromosome. The resultant loss of the plasmid would allow the bacterium to regain its fitness whilst keeping any beneficial traits (Harrison and Brockhurst 2012). Over the last decade, several solutions to this paradox have been shown to operate in laboratory bacterial communities and populations (Brockhurst and Harrison 2022). These include Infectious transmission; Variable host background; Plasmid interactions; Source sink transmission; Piggybacking adaptions and compensatory evolution see (**figure 3.2**) for brief explanation.

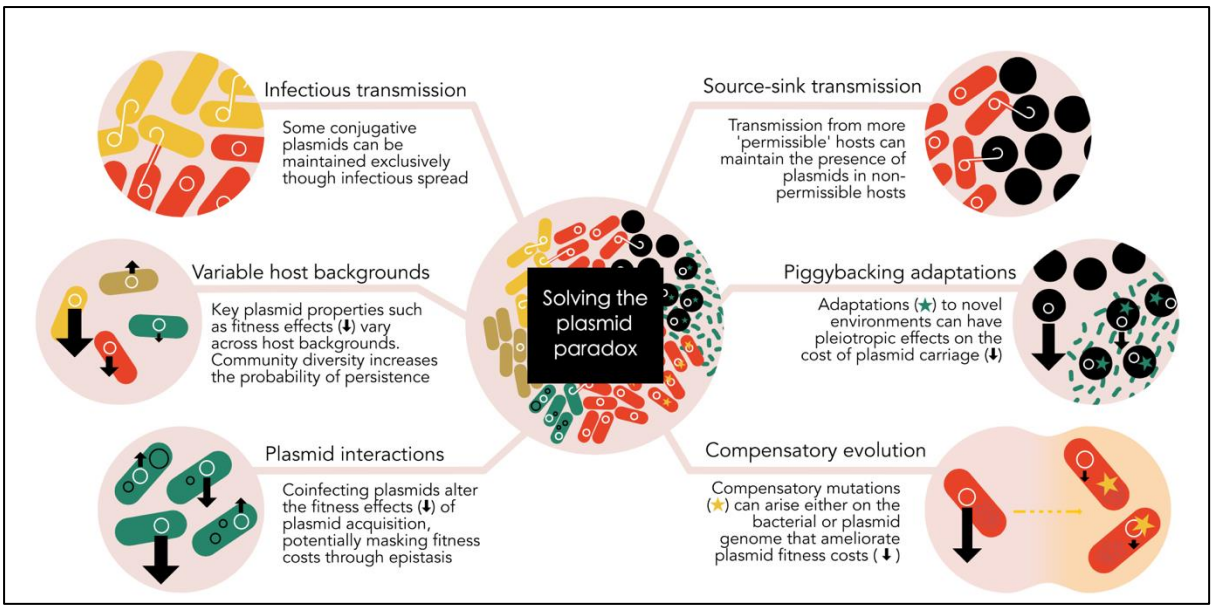


Figure 3.2. Ecological and Evolutionary mechanisms of Plasmid stability.
Over the last decade, theoretical and experimental studies have discovered a range of ecological and evolutionary mechanisms that enable plasmids to persist in bacterial populations and communities. Cells of the same colour belong to the same species. Whereas those of different colour belong to a different species or lineage. Plasmids are shown as white rings, and where these connect cells, this denotes conjugative transfer. Solid arrows denote fitness costs and are scaled to the magnitude of the fitness cost. Stars indicate mutations, with the specific type of mutation described in the figure. Broken lines denote evolutionary changes.

However, how the plasmid paradox plays out in real world situations is largely unknown.

3.1.4 Chromosomal carriage of antibiotic resistance genes

There are limited numbers of publications describing a chromosomal location for antibiotic genes in the literature (**Table 3.1**). Most of these studies only involve small numbers of isolates and exact chromosomal insertion sites and size of the insertions are often not identified. Isolates are investigated for chromosomal location by I-Ceu digestion of genomic DNA and probing of Southern blots or by iPCR or combinations of short and long-read sequencing. I Ceu-1 analysis is based on the fact, that this rare cutting restriction enzyme only digests chromosomal DNA at RNA operons and spares plasmids. Southern blots of PFGE gels are then probed with an RNA probe as well as the relevant resistance gene. Since RNA operons are only encoded chromosomally, where large fragments are positive for both probes the gene is assigned a chromosomal location. Sequence information is used in the other

techniques to determine that resistance genes are adjacent to known chromosomal genes.

3.1.4.1 Chromosomal location of antibiotic resistance genes in *Proteus mirabilis*

Several of the studies have described common chromosomal carriage of antibiotic resistance genes in *Proteus mirabilis* ranging from 75-100% of the isolates investigated (**Table 3.1**). These include 100% chromosomal location of *bla*$_{\text{OXA-23}}$ and *bla*$_{\text{VIM-1}}$ genes in several *P. mirabilis* clinical isolates from individual institutions in France and Greece, respectively (Bonnet et al. 2002; Tsakris et al. 2007). National studies in Korea and Japan also identified 75% chromosomal carriage of *bla*$_{\text{CTX-M}}$ genes (21/28) and 80% chromosomal carriage of *bla*$_{\text{CTX-M-2}}$ (4/5) in *P. mirabilis* clinical isolates from Korea and Japan, respectively (Song et al. 2011; Harada et al. 2012). Lastly, an international study identified 100% chromosomal carriage of *bla*$_{\text{CMY}}$ genes in *P. mirabilis* isolates collected from France, Greece, Italy and Poland (D'Andrea et al. 2011). Of these studies, only the last one identifies the chromosomal insertion site and size of inserts. All *P. mirablis* isolates had an identical size and site of insertion. This was an IS*Ec*p1 mediated insertion of 6,210 bp including the *bla*$_{\text{CMY}}$ gene and adjacent chromosomal genes *blc* and *sug*E derived from *Citrobacter freundii* as well as a section of a *Col*E1-type plasmid (*orf6*, *mob* and RNAII). Several isolates also had a second insertion at a different chromosomal site.

3.1.4.2 Chromosomal location of antibiotic resistance genes in *E. coli*

Several papers have described a chromosomal location for resistance genes in *E. coli*. These include a study of 66 CTX-M-9 producing *E. coli's* from the Ramon y Cajal hospital in Madrid from 1996-2003 where only a single *E. coli* was found with

bla$_{CTX-M-9}$ in a chromosomal location 1.5% (1/66) (Novais et al. 2006). Coque *et al* 2008 studied 43 CTX-M-15 producing clinical *E. coli* from France, Kuwait, Switzerland, Canada, Portugal, and Spain (2000-2006) that were representative of these areas because they either caused outbreaks or were the first isolates recovered in those countries. They reported a chromosomal location for 18.6% (8/43)(Coque et al. 2008). A similar chromosomal carriage rate, 22% (40/185) was also found for CTX-M positive UTI *E. coli* isolates collected in 2017 from Bristol in the UK. These were found in 11 STs with STs 131 (n=11) and 73 (n=10) being the most represented. Whilst in ST131 the *bla*$_{CTX-M}$ genetic environments were diverse, in ST73, 9/10 isolates harboured the gene in the same genomic location, suggesting a high degree of clonality within this ST. However, a Korean study of CTX-M-14 producing *E. coli's* isolated from 11 hospitals in 2005 found a higher carriage rate of chromosomal carriage of 37% (13/35). Interestingly, Rodriguez *et al* 2014 studied 356 *E. coli* Isolates collected from animals, animal products and humans in the Netherlands, the UK and Germany 2005-2009 that were resistant to 3rd generation cephalosporins as part of the SAFEFOODERA-ESBL project. They found chromosomally encoded CTX-M genes in only 4.8% (17/356) of isolates. Notably, these were only from human sources which were mainly clinical isolates and represented 12.9% (17/132) of all human isolates (Rodríguez et al. 2014). A few studies have further information on insertion sites using either inverse PCR (iPCR) methodologies or information gained from short read sequencing. Hirai *et al* 2013 used an iPCR technique to identify chromosomal insertion sites in 21 *E. coli* 025b-B2-ST131 isolates collected from university hospital patients and nursing care home residents in the Kinki region of Japan. They found a collection of IS*Ec*p1 mediated insertions of an identical length (2971 bp) that included IS*Ec*P1 and *bla*$_{CTX-M-15}$ in

eight different insertion sites. These sites were identified as intergenic and within genes encoding; choline dehydrogenase (*betA*), d-serine dehydratase (*dsdA*), d-serine deaminase activator (*dsdC*), d-serine permease (*dsdX*), ferrichrome-iron receptor (*fhuA*), flagellar transcriptional activator protein (*flhC*), transcriptional activator protein (*flhD*), trehalose-6-phosphate synthase (*otsA*), ATP phosphoenolpyruvate carboxykinase (*pck*) and hypothetical protein YhgE (*yhgE*). Five of these insertions included a target site duplication characteristic of transposition by IS*Ec*P1 and 13 isolates were in an identical position in *dsd*A with PFGE indicating similar but not identical patterns (Hirai et al. 2013).

Four publications used long-read sequencing to identify chromosomal insertion sites. One of these have a limited number of isolates with little reference to national or international relevance. Decano *et al* 2019 investigated 6 ESBL *E. coli* ST131 isolates from six patients at Addenbrokes hospital, Cambridge in 2015 and identified chromosomal $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-14}}$ in two isolates. Interestingly, this study recorded three separate chromosomal insertions of $bla_{\text{CTX-M-15}}$ in a single isolate. All insertions were of short lengths and included IS*Ec*P1, the $bla_{\text{CTX-M}}$ gene and a short section of Tn3. Additional information of insertion sites were not described (Decano et al. 2019). Goswami *et al* 2020 investigated 16 *E. coli* bloodstream isolates that were representative of a larger collection of 162 isolates collected 2013-15. Seven of these contained $bla_{\text{CTX-M-15}}$, 5 of which were on the chromosome. They used a combination of short and long reads to assemble complete genomes and note that three chromosomal insertions included the $bla_{\text{OXA-1}}$ gene. They also note the additional importance of IS26 in these mobilisation events but give no further detail of insertion site (Goswami et al. 2020). Zahra *et al* investigated 110 *E. coli* that were representative of the entire carriage types of *E. coli* (without antibiotic bias) within the

population of Islamabad in Pakistan in 2017. The ESBL gene *bla*$_{\text{CTX-M-15}}$ was found in 31% of all *E. coli* and 23% of isolates carried it in a chromosomal location. In all chromosomally encoded isolates, no plasmid copies of *bla*$_{\text{CTX-M-15}}$ were found suggesting purifying selection characteristic of the plasmid paradox. Four insertions were investigated in depth and consisted of a 11,389 bp insertion found in three isolates of ST394 including *bla*$_{\text{CTX-M-15}}$ and *qnrS*1 and was identical to the first 11 389 bp of an isolate of ST4121 that included additional *tetA* and *tetR* resistance genes as well as several other genes of plasmidic origin. The chromosomal insertion in the ST394 isolates was in a low GC% intergenic region between *ydcS* and *ydcR* genes encoding a putative DNA-binding transcriptional regulator and a putative spermidine/putrescine transporter subunit, respectively. The insertion in the ST4121 isolate was in a hydrolase gene. Each insertion had the mobile element IS*Ec*p1 at the left-hand terminus and was flanked by target-site direct repeats of TATGA for ST4121 and TTAAA for isolates ST394 (Zahra et al. 2018).

Large chromosomal insertions including several antibiotic-resistance genes were also found in a recent study of Zambian isolates at a single institution in 2018 (Shawa et al. 2021). They found a chromosomal insertion rate of 15.5% (7/45) for clinical Enterobacterales. This included 17% (6/35) for *E. coli* and single *Enterobacter cloacae.* These insertions included *bla*$_{\text{OXA-1}}$, ***aac6'-ib-cr*** and *cat* genes in addition to *bla*$_{\text{CTX-M-15}}$. Chromosomal locations have seldom been described for antibiotic resistance genes in other Enterobacterales (**Table 3.1**).

Table 3.1 Chromosomal location papers

Paper	link	isolates studied	Detail
References reporting a chromosomal genetic location of antibiotic resistance genes in *Proteus mirabilis*			
Bonnet *et al* 2002	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC127228/pdf/0704.pdf	Ten *Proteus mirabilis* isolates were collected at a single institution in Clermont-Ferrand, France from 1996-1999. PFGE indicate 9/10 were clonal and the remaining one closely similar. All produced OXA-23	100% Chromosomal location was identified by the I-CeuI method. No information on chromosomal site or size of insertion.
Tsakris *et al* 2007	https://academic.oup.com/jac/article/60/1/136/730295	12 MBL-positive *Proteus mirabilis* isolates were recovered from urinary tract infections 2005-2006 of community patients at a single institution in Greece. In all cases, the patients had a previous hospitalization in a Greek regional hospital.	100% chromosomal location determined by I-CeuI method. Insertion site and size were not investigated.
Song *et al* 2011	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3067170/	Korean nationwide survey conducted between July and September 2008, a total of 222 consecutive nonduplicate *Proteus mirabilis* isolates were collected from 17 hospitals in 11 cities in Korea. The isolates were recovered from urine (138), blood (10), respiratory specimens (18), and pus (56). 28 (12.6%) and 8 (3.6%) isolates exhibited extended-spectrum β-lactamase (ESBL) and AmpC phenotypes.	$bla_{\text{CTX-M}}$ genes were located on the chromosome in 21/28 ESBL (75%) isolates. There were 2 isolates with CTX- M-2, 3 with CTX-M-12, 9 with CTX-M-14, 3 with CTX-M- 15, and 4 with CTX-M-90. Furthermore, probes specific for the $bla_{\text{DHA-1}}$ and bla_{CMY-2} genes also hybridized with 7/8 ampC isolates, two and five isolates, respectively (88%).
D'Andrea *et al* 2011	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3101460/	21 *Proteus mirabilis* strains from 4 countries France Italy Poland and Greece collected 1999-2008 from various clinical sources. All isolates	All isolates had a chromosomal location by the I-CeuI methodology and all isolates had an ISEcP1 allele 110 bp upstream of the gene cmy-4 or closely related variant. Size of inserts were identified as well as insertion site by cloning, PCR mapping and sequencing. All insertions

Paper	link	isolates studied	Detail
		harboured an ampC type resistance mechanism and had different PFGE profiles.	were in the pepQ gene encoding a Xaa-Pro dipeptidase and insertions were of 6210bp and named Tn6093 with same position and same direct repeat of the gene. Tn*6093* contained IS*Ecp1*, followed by the *C. freundii*-derived *bla*$_{CMY}$, *blc*, and *sugE* genes, a part of a ColE1-type plasmid backbone (regions *orf6*, *mobB*, and RNAII), and terminating with an IS*Ecp1* IRR-like sequence, named IRR-4. In isolates with two *bla*$_{CMY}$ copies, a second IS*Ecp1*- *bla*$_{CMY}$ module, 3,783 bp long and named Tn*6113*, was inserted into the *P. mirabilis* intergenic spacer between ORF PMI0120 and the *ppiD* gene encoding peptidyl-prolyl *cis-trans* isomerase D. In all of these isolates, Tn*6113* was inserted at the same position, 154 bp downstream of *ppiD*, and was flanked by 5-bp direct repeats (TATGA), which indicated that this module also had integrated into the chromosome by transposition. The data is consistent with a single mobilization of the *bla*$_{CMY}$ ancestor and its putative donor to the *P. mirabilis* chromosome. This was most probably a ColE-like plasmid similar to pTKH11, that carries an IS*Ecp1* module with *bla*$_{CMY}$ - like genes, named Tn*6114*.
Harada *et al* 2012	https://journals.asm.org/doi/epub/10.1128/AAC.00258-11	76 clinical *Proteus mirabilis* isolates collected during a Nationwide surveillance study from 11 health care facilities in Japan between September 2005 and March 2006. Five of them (6.5%) were ESBL producers all ESBLs were CTX-M-2. Two from one site and all others from geographically different sites.	A chromosomal location was determined in 80% (4/5) isolates by the I-CeuI technique including two that also harboured a plasmid copy and one isolate with a unique plasmid location. ISEcp1 was responsible for integration in all chromosomal isolates and included three separate insertion sites one of which encoding an acetyl coenzyme synthetase gene and others into unknown ORFS. One strain harboured two insertions on the chromosome. They used inverse PCR to locate the insertion site. All inserts were of about 3kb with most insertions having direct repeats of 5bp giving hallmarks of transposition.
References reporting a chromosomal genetic location of antibiotic resistance genes in *Escherichia coli*			

Paper	link	isolates studied	Detail
Novais *et al* 2006	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1538643/	Seventy CTX-M-9-producing clinical isolates (66 *Escherichia coli* isolates, 2 *Klebsiella pneumoniae* isolates, 1 *Enterobacter cloacae* isolate, and 1 *Salmonella enterica* isolate) collected from 45 patients at Ramón y Cajal Hospital, a 1,200-bed university teaching hospital in northern Madrid, Spain (1996 to 2003)	Chromosomal location was found in a single *E. coli* isolate 1/66 (1.5%) No information on insertion size or site is given.
Coque *et al* 2008 (Note ref 4 from Rodriquez paper)	https://wwwnc.cdc.gov/eid/article/14/2/07-0350_article	43 CTX-M-15–producing *E*. *coli* clinical isolates from France (17), Kuwait (9), Switzerland (7), Canada (4), Portugal (3) and Spain (3), and 6 CTX-M-15 plasmids from India (*3*), all obtained from 2000 through 2006. These strains and plasmids were considered representative of these areas because they either caused outbreaks or were the first isolates recovered in those countries. Samples were isolated from urine (n = 33/43, 77%), wounds (n = 4/43, 9%), respiratory tract infections (n = 3/43, 7%) and other sites (1 from feces, 1 from an intravenous catheter, and 1 from blood) in hospitalized patients.	Detected a chromosomal location in 8/43 (18.6%) of isolates by probing of I-CeuI PFGE gels. (2 belonging to B2 (ST131), 2 to D (ST405+unknown), and 1 to A. In 2 other strains, the probe hybridized with plasmid and chromosomal bands (1 strain from D ST405 and 1 from phylogroup B1). Paper gives no information of insertion size or site but notes the close association of CTX-M-15 with OXA-1 and aac6'-ib'cr. No information on genetic elements responsible.
Kim *et al* 2011 (Note reference 5 from Rodriquez paper)	https://academic.oup.com/jac/article/66/6/1263/718268	138 non-duplicate *E. coli* isolates isolated in 11 different hospitals in Korea between May and June 2005. They all showed reduced susceptibility or resistance to ceftazidime and/or cefotaxime. Of these 35 encoded the $bla_{\text{CTX-M-14}}$ gene.	The $bla_{\text{CTX-M-14}}$ gene was located on the chromosome in 13/35 (37%) *E. coli* clinical isolates in this study. The gene was surrounded by genetic environments like those on plasmids, supporting the hypothesis that the $bla_{\text{CTX-M-14}}$ gene integrated into the chromosome via transposable elements such as ISEcp1. This was just by PCR so insertion site and size of insertion was not determined.

Paper	link	isolates studied	Detail
Hirai *et al* 2013	https://pubmed.ncbi.nlm.nih.gov/24091130/	22 *E. coli* O25b-B2-ST131 strains 21/22 harbouring chromosomal *bla*$_{CTX-M-15}$ isolated from university hospital patients and nursing home residents in the Kinki region of Japan.	Identified chromosomal insertions by an iPCR method giving limited information about insertion and insertion sites. Identified many different insertions and mention no evidence of clonal spread. Each insertion was identical length of 2971 bp and collection consisted of 8 different insertion sites A-H. 5 of these included a target site duplication A & E-H. The identified genes are indicated as choline dehydrogenase (betA), d-serine dehydratase (dsdA), d-serine deaminase activator (dsdC), d-serine permease (dsdX), ferrichrome-iron receptor (fhuA), flagel- lar transcriptional activator protein (flhC), transcriptional activator protein (flhD), trehalose-6-phosphate synthase (otsA), ATP phos- phoenolpyruvate carboxykinase (pck) and hypothetical protein YhgE (yhgE). 13 isolates harboured the type A insertion into the dsdA gene but PFGE indicated similar but not identical patterns The prevalence of *E. coli* isolates possessing chromosomal bla$_{CTX-M-15}$ was only 0.3% in 2008 at the university hospital and 8.3% among nursing home residents in 2010.
Rodriguez *et al* 2014	dx.doi.org/10.1016/j.ijantimicag.2014.02.019	356 (subset of larger group 629) Isolates collected from animals, animal products and humans in the Netherlands, the UK and Germany 2005-2009 resistant to 3rd generation cephalosporins as part of SAFEFOODERA-ESBL project.	Detected chromosomal ESBL in 17/356 (4.8%) isolates due to inability to transfer resistance in conjugation experiments. These all were from human sources 17/132 (12.9%) and mostly hospital related. Notably chromosome location was not found from animal sources (0/224). Isolates belonged to phylogenetic groups B1 (2), B2 (6) and D (9) and 7 sequence types: ST38 (7), ST131 (4), ST648 (2) and singletons of ST156, ST1266, ST2178, and ST3878. ESBLs were CTX-M15 (11), CTX-M-14 (4), CTX-M-9(1) CTX-M-51(1). The insertions and insertion sites were not studied. They suggest that two strains carry more than one CTX-M gene (probing I-CeuI gels).
Zahra *et al* 2018	https://pubmed.ncbi.nlm.nih.gov/29648612/	Investigated 110 *E. coli* that were isolated from numerous sewage sites across Islamabad, Pakistan and represented the common human	*bla*$_{CTX-M-15}$ was found in 31% of all *E. coli* and 23% of isolates carried it in a chromosomal location. In all chromosomally encoded isolates no

Paper	link	isolates studied	Detail
		carriage types in the city in 2017, without any antibiotic bias.	plasmid copies of *bla*$_{CTX-M-15}$ were found. Four insertions were investigated in depth and consisted of a 11,389 bp insertion found in three isolates of ST394 including *bla*$_{CTX-M-15}$ and *qnrS*1 and was identical to the first 11 389 bp of the 23,174 bp chromosomal insertion of isolate of ST4121 that included additional tet(A) and tet(R) resistance genes as well as several other genes of plasmid origin. The chromosomal insertion in the ST394 isolates was in a low GC% intergenic region between ydcS and ydcR genes encoding a putative DNA- binding transcriptional regulator and a putative spermidine/putrescine transporter subunit, respectively. The insertion in the ST4121 isolate was in a hydrolase gene. Each insertion had the mobile element ISEcp1 at the left-hand terminus and was flanked by target-site direct repeats of TATGA for ST4121 and TTAAA for isolates ST394.
Decano et al 2019	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6506616/	6 ESBL *E. coli* ST131 isolates from six patients. *bla*$_{CTX-M-15}$, *bla*$_{CTX-M-14}$, and *bla*$_{CTX-M-27}$ genes identified in three, one, and one isolates, respectively. One sample had no *bla*$_{CTX-M}$ gene.	Two samples had chromosomal *bla*$_{CTX-M-14}$ and *bla*$_{CTX-M-15}$ genes, and the latter was at three distinct locations, likely transposed by the adjacent MGEs: IS*Ecp1*, IS*903B,* and Tn*2*. Methodology included long read sequencing.
Findlay *et al* 2020		900 Cefalexin-resistant UTI *E. coli* isolates from 836 patients collected during routine urine microbiology testing at Severn Infection Partnership, Southmead Hospital between September 2017 and August 2018. These were from 146 general practices located throughout Bristol and including coverage in Gloucestershire, Somerset and Wiltshire. Almost 69% (576/ 836) were cefotaxime resistant, Ciprofloxacin resistance was observed in 424/836 (50.7%) of all	Forty *bla*$_{CTX-M}$ genes, of variants *bla*$_{CTX-M-14}$ (n=3) and *bla*$_{CTX-M-15}$ (n=37), were found to be located on the chromosome 40/185 (22%) By analysis of short read sequence data. These were found in 11 STs with STs 131 (n=11) and 73 (n=10) being the most represented. Whilst in ST131 the *bla*$_{CTX-M}$ genetic environments were diverse, in ST73, 9/10 isolates harboured the gene in the same genomic location, suggesting a high degree of clonality within this ST. No information given on the chromosomal sites or size of insert. Also suggested OXA gene on the chromosome in ST38 isolate.

Paper	link	isolates studied	Detail
		isolates and in 363/576 (63.0%) of cefotaxime-resistant isolates. Two-hundred and twenty-five isolates, chosen to be representative of resistance gene carriage and patient demographics (age, sex) were selected for WGS. The most common cefotaxime-resistance mechanism was CTX carriage (185/ 225), followed by plasmid-mediated AmpCs (pAmpCs) (17/225), AmpC hyperproduction (13/225), ESBL SHV variants (6/225) or a combination of both CTXM and pAmpC (4/225).	
Goswami 2020	https://pubmed.ncbi.nlm.nih.gov/32160146/	16 *E. coli* bloodstream isolates from Scotland collected 2013-15. These were representative of a larger collection of 162 isolates. 12 were ST131 and 4 were ST38 isolates. From these they assembled 46 plasmids. 7 strains contained CTX-M-15	5/7 strains had the *bla*$_{\text{CTX-M-15}}$ gene on the chromosome. They used a combination of short and long read sequencing to assembly complete genome and the chromosomal location and size was mapped for 5 chromosomal locations. No details of insertion are given in the paper other than the diagrams including insertion near yeeZ etc of the genes associated (OXA is associated with three of the chromosomal integrations) and they note the importance of IS26 in the insertion event but not detail on direct repeats etc.
Shawa *et al* 2021	https://aricjournal.biomedcentral.com/articles/10.1186/s13756-021-00941-8	46 clinical isolates of cefotaxime-resistant Enterobacteriaceae (1 *Enterobacter cloacae*, 9 *Klebsiella pneumoniae*, and 36 *Escherichia coli*) from Zambia were subjected to whole-genome sequencing (WGS) using MiSeq and MinION.	Out of 45 *bla*$_{\text{CTX-M}}$ gene-carrying strains, 7 (15.5%) harbored the gene in their chromosome. In one *E. cloacae* and three *E. coli* strains, chromosomal *bla*$_{\text{CTX-M-15}}$ was located on insertions longer than 10 kb. These insertions were bounded by IS*Ec*p1 at one end, exhibited a high degree of nucleotide sequence homology with previously reported plasmids, and carried multiple AMR genes that corresponded with phenotypic AMR profiles.
References reporting a chromosomal genetic location of antibiotic resistance genes in other Enterobacterales			

Paper	link	isolates studied	Detail
Coelho *et al* 2010	https://pubmed.ncbi.nlm.nih.gov/20392607/	37 CTX-M-15-producing *Klebsiella pneumoniae* isolates collected from 2005 to 2008 within the Barcelona metropolitan area The 37 studied *K. pneumoniae* isolates represented all the CTX-M-15-producing isolates found among 111 ESBL-producing *K. pneumoniae* collected from three hospitals within metropolitan Barcelona between 2005 and 2008	The majority of the 37 isolates carried $bla_{CTX\text{-}M\text{-}15}$ in a plasmid location, frequently associated with the aac(6)-Ib-cr gene. Plasmids encoding $bla_{CTX\text{-}M\text{-}15}$ carried three distinct replicons, i.e. IncFII, IncR and IncFIIk, the latter two not having been described previously in association with $bla_{CTX\text{-}M\text{-}15}$. Several of these plasmids were not self-transferable. Furthermore, in all isolates belonging to sequence type ST-1 32% (12/37), $bla_{CTX\text{-}M\text{-}15}$ was found integrated into the *K. pneumoniae* chromosome. In all the studied isolates, the mobile element ISEcp1 was found upstream of $bla_{CTX\text{-}M\text{-}15}$ whereas IS26 was found inserted within ISEcp1 in several isolates.
Fabre *et al* 2009	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2681507/	*Salmonella enterica* serotype concord isolates from France and Norway isolated 2001-2006 including 43 from France and 26 from Norway.	A subset of 12 CTX-M-15 isolates were investigated for genomic position by transfer experiments 8/12 failed to transfer suggesting a chromosomal location. Analysis of integration site by a cloning methodology combined with primer extension gave a partial analysis of the chromosomal insertion and included sections of incAC plasmid previously found to harbour CTX-M-15 in this species.

Here we used a large contemporary collection of *E. coli* isolates (collected August 2013- December 2014) in five regions of the UK to investigate the role played by chromosomal insertion of ESBL genes originating from different sources. Isolates were collected from bacteraemias and faecal samples from Hospitals located in London, Wales, Scotland, East Anglia and the North-West. At the same time, ESBL *E. coli* were also sourced from sewage works, animal slurry and supermarket meat in these regions. As far as we are aware this represents the largest contemporary collection of ESBL-*E. coli* investigated for a chromosomal location anywhere in the world and an ideal starting point to determine the importance of chromosomal carriage of antibiotic resistance genes in UK *E. coli.*

3.2 Methodological rationale

Faecal, bacteraemia and animal ESBL *E. coli* isolates were investigated for chromosomal location by in-gel probing of S1 nuclease partially digested *E. coli* pulsed-field gel plugs (Toleman 2018). S1 nuclease treatment of PFGE plugs essentially nicks plasmid DNA so that the plasmids take an open circular and/or linear form and run true to size when separated by pulsed-field gel electrophoresis (PFGE). This is due to the fact, that S1 nuclease has a bifunctional activity being both a double-stranded DNA exonuclease and a single-stranded DNA endonuclease. Since the single-stranded DNA endonuclease activity is more potent than its exonuclease activity, dilute concentrations of the enzyme are used to nick exposed single-stranded DNA in supercoiled coiled plasmids. Moreover, the chromosome is not digested so there is a clear distinction between the chromosome and plasmids when separated by size during PFGE. When combined with in-gel probing, S1 digestion is superior to the I-Ceu-1 methodology as it only requires probing with the antibiotic-resistance gene of interest. This is also because very

large pieces of DNA do not transfer well to the nylon membrane during the Southern blot procedure and give faint or even misleading results during the I-Ceu-I or S1 methodologies. The in-gel hybridisation renders Southern blotting unnecessary yielding excellent results. Isolates that had a chromosomal location were further investigated for a subset of isolates (Toleman et al. 2020).

By investigation of short-read sequence information and subsequently a larger group by long-read sequencing using minion flow cells and minion sequencer. Exact locations were investigated by in silico analysis using geneious software and the chromosomal insert size was determined as described in detail in chapter 2. The chromosomal location was then drawn to scale.

3.2.1 Detail of minion long read sequencing and assemblies.

The isolates were sequenced using several Nanopore minion flow cells. Each DNA was individually barcoded using the SQK-RBK 110.96 barcoding kit before adding the pooled and barcoded DNA to each flowcell. Minion flowcells were run for 72 hrs and barcoded FAST5 files were loaded onto the Hawk supercomputer in Cardiff and translated into FASTQ files using Guppy software. The FASTQ files were then assembled using Flye software and annotated using Prokka. Details of isolates/ flowcells/ number of reads and assembly statistics are given in (**Table 3.2**) below. Assemblies were of various qualities which was mostly due to the depth of coverage. Several isolates were completely closed, and chromosomal location was identified using local blast searches using Geneious software (Biomatters. USA).

Table 3.2 Detail of Nanopore minion flowcells/isolate ID and assembly statistics

NANOPORE run	Isolate/ Barcode	#FAST5 folders	#FASTQ folders	Miseq?	Flye assembly (#contigs)	Length	Coverage
	RAHMAN[R1]						
FAS43138	DH212/ BC27	7	7	N	Y (7)	5,096,420	28 X
FAS43138	DH231/BC31	19	19	N	Y (8)	5,315,008	58 X
FAS43138	DH356/ BC18	8	8	N	Y (14)	5,168,095	36 X
FAS43138	DH458/BC24	14	14	N	Y (10)	5,330,639	43 X
FAS43138	DH784/BC17	13	13	N	Y (17)	5,435,699	35 X
FAS43138	DH1540/BC22	11	11	N	Y (43)	5,220,746	24 X
FAS43138	DH1540b/BC35	16	16	N	Y (5)	5,340,395	58 X
FAS43138	DH1684/BC40	12	12	N	Y (14)	5,250,990	44 X
FAS43138	DH1705/BC39	18	18	N	Y (12)	5,416,617	55 X
FAS43138	DH1709/BC42	7	7	N	Y (14)	5,468,461	20 X
FAS43138	DH2202/BC10	3	3	N	Y (76)	4,307,367	8 X
FAS43138	DH2205/BC25	12	12	N	Y (31)	5,136,750	27 X
FAS43138	DH2206/BC23	13	13	N	Y (43)	5,220,746	24 X
FAS43138	DH2218/BC16	47	47	N	Y (5)	5,370,617	201 X
FAS43138	DH2220/BC12	13	13	N	Y (25)	5,191,135	43 X
FAS43138	DH2221/BC29	6	6	N	Y (2)	5,408,573	23 X
FAS43138	DH2222/BC19	12	12	N	Y (3)	5,474,688	46 X
FAS43138	DH2223/BC28	12	12	N	Y (3)	5,376,272	28 X
FAS43138	DH2229/BC34	20	20	N	Y (8)	5,361,701	66 X
FAS43138	DH2230/BC14	9	13	N	Y (26)	5,197,545	35 X
FAS43138	DH2238/BC11	8	8	N	Y (25)	5,180,898	25 X
FAS43138	DH2240/BC21	8	8	N	Y (34)	5,210,123	25 X
FAS49725	1/DH1707	20	20	N	Y (4)	4,653,928	103 X
FAS49725	2/DH1105	14	14	N	Y (3)	4,995,414	70 X
FAS49725	3/DH1090	5	5	N	Y (6)	5,205,626	14 X
FAS49725	4/DH1098	4	4	N	Y (15)	5,458,179	11 X
FAS49725	5/DH1703	8	8	N	Y (4)	5,380,270	37 X
FAS49725	6/DH1099	16	16	N	Y (4)	4,964,114	57 X
FAS49725	7/DH1689	16	16	N	Y (4)	5,393,121	36 X
FAS49725	8/DH1702	10	10	N	Y (2)	5,328,119	33 X
FAS49725	9/DH1687	20	20	N	Y (2)	5,192,223	70 X
FAS49725	10/DH1710	17	17	N	Y (7)	5,461,530	64 X
FAS49725	11/DH1696	9	9	N	Y (2)	5,187,309	39 X
FAS49725	12/DH1547	8	8	N	Y (6)	5,381,457	29 X
FAS49725	13/DH1543	17	17	N	Y (8)	5,199,564	49 X
FAS49725	14/DH1647	11	9	N	Y (8)	5,477,792	38 X
FAS49725	15/DH1545	9	9	N	Y (1)	4,703,214	32 X
FAS49725	16/DH1704	2	2	N	Y (72)	4,529,594	5 X
FAS49725	17/DH1700	13	13	N	Y (7)	5,278,217	31 X
FAS49725	18/DH1680	18	18	N	Y (2)	5,382,480	47 X
FAS49725	19/DH1535	26	26	N	FAIL		
FAS49654	DH107/BC1	10	10		Y (18)	5,387,358	28 X
FAS49654	DH677/BC2	18	18		Y (5)	5,288,698	60 X
FAS49654	DH2227/BC3	6	6		Y (33)	5,292,513	14 X
FAS49654	DH793/BC5	20	20		Y (7)	5,361,219	64 X
FAS49654	DH785/BC6	5	5		Y (45)	4,926,177	10 X
FAS49654	DH782/BC7	8	8		Y (4)	5,333,892	22 X

NANOPORE run	Isolate/ Barcode	#FAST5 folders	#FASTQ folders	Miseq?	Flye assembly (#contigs)	Length	Coverage
FAS49654	DH776/BC8	8	8		Y (32)	5,104,014	17 X
FAS49654	DH104/BC10	12	12		Y (13)	5,354,598	29 X
FAS49654	DH902/BC11	5	5		Y (70)	4,694,530	9 X
FAS49654	DH246/BC12	25	25		Y (11)	5,236,997	52 X
FAS49654	DH810/BC13	8	8		Y (11)	5,005,436	22 X
FAS49654	DH826/BC15	9	9		Y (19)	5,372,005	22 X
FAS49654	DH935/BC31	10	10		Y (13)	5,602,439	33 X
FAS49654	DH108/BC33	9	9		Y (35)	5,217,043	16 X
FAS49654	DH777/BC34	7	7		Y (27)	5,177,149	16 X
FAS49654	DH1134/BC35	12	12		Y (9)	5,123,026	34 X
FAS49654	DH739/BC37	9	9		Y (16)	5,080,178	24 X
FAS49654	DH761/BC38	13	13		Y (9)	5,343,748	36 X
FAS49654	DH727/BC40	3	3		Y (81)	4,235,155	6 X
FAS49654	DH105/BC41	15	15		Y (19)	5,340,799	39 X
FAS49654	DH59/BC42	2	2		Y (52)	5,049,354	6 X
FAS49654	DH734/BC43	2	2		Y (7)	5,152,947	11 X
FAS49654	DH228/BC44	8	8		Y (2)	5,224,888	27 X
FAS49654	DH234/BC45	17	17		Y (5)	5,438,831	56 X
FAS49654	DH754/BC46	3	3		Y (45)	4,735,355	9 X
FAS49654	DH1191/BC47	5	5		Y (22)	4,989,427	17 X
FAS49654	DH115/BC48	2	2		Y (19)	5,165,901	9 X
FAS49654	DH804/BC50	6	6		Y (24)	5,321,525	13 X
FAS49654	DH941/BC51	6	6		Y (26)	4,807,352	14 X
FAS49654	DH112/BC52	10	10		Y (14)	5,218,626	20 X
FAS49654	DH897/BC53	6	6		Y (20)	5,125,673	13 X
FAS49654	DH923/BC54	5	5		Y (35)	5,389,379	12 X
FAS49654	DH726/BC55	17	17		Y (12)	5,220,493	31 X
FAS49654	DH763/BC56	18	18		Y (6)	5,196,448	43 X
FAS49654	DH839/BC57	5	5		Y (11)	5,320,028	14 X
FAS49654	DH925/BC58	4	4		Y (42)	5,006,337	9 X
FAS49654	DH834/BC59	6	6		Y (42)	5,006,337	9 X
FAS49654	DH1201/BC60	4	4		Y (54)	4,550,706	9 X
FAS49654	DH878/BC61	1	1		Y (46)	3,293,742	4 X
FAS49654	DH520/BC63	11	11		Y (7)	5,425,726	32 X
FAS49654	DH712/BC65	8	8		Y (41)	5,222,678	14 X
FAS49654	DH727/BC66	6	6		Y (54)	5,045,879	10 X
FAS49654	DH208/BC67	3	3		Y (58)	4,776,951	6 X
FAS49654	DH551/BC68	7	7		Y (20)	5,188,265	15 X
FAS49654	DH694/BC69	6	6		Y (29)	5,259,540	11 X
FAS49654	DH240/BC70	7	7		Y (22)	4,824,617	14 X
FAS49654	DH536/BC71	6	6		Y (32)	4,712,120	11 X
FAS49654	DH964/BC72	12	12		Y (8)	4,883,701	25 X
FAS49654	DH927/BC81	6	6		Y (37)	5,358,877	37 X
FAS49654	DH506/BC82	4	4		Y (5)	5,287,748	13 X
FAS49654	DH233/BC83	19	19		Y (2)	5,384,311	12 X
FAS49654	DH722/BC84	7	7		Y (20)	5,306,070	16 X
FAS49654	DH719/BC85	9	9		Y (6)	5,182,050	25 X
FAS49654	DH716/BC86	4	4		Y (9)	5,508,540	12 X
FAS49654	DH95/BC87	17	17		Y (14)	5,504,119	43 X
FAS49654	DH1688/BC88	6	6		Y (25)	5,323,041	16 X
FAS49654	DH1215/BC89	3	3		Y (17)	5,295,142	9 X

NANOPORE run	Isolate/ Barcode	#FAST5 folders	#FASTQ folders	Miseq?	Flye assembly (#contigs)	Length	Coverage
FAS49654	DH951/BC90	4	4		Y (17)	4,957,052	11 X
FAS49654	DH857/BC91	4	4		Y (13)	5,024,434	11 X
FAS49654	DH976/BC92	6	6		Y (26)	5,292,734	13 X
FAS49654	DH806/BC94	5	5		Y (18)	5,455,484	11 X
FAS49654	DH803/BC95	6	6		Y (19)	5,588,440	14 X
FAS43138	DH80/BC15	30	30	N	Y (7)	5,246,084	103 X
FAS43138	DH766/BC37	11	11	N	Y (17)	5,231,660	34 X
FAS43138	DH795/BC36	4	4	N	Y (65)	5,080,169	9 X
FAS43138	DH2197/BC41	3	3	N	Y (7)	5,179,090	9 X
FAS43138	DH2203/BC32	12	12	N	Y (21)	5,277,808	34 X
FAS43138	DH2210/BC30	7	7	N	Y (16)	5,293,366	23 X
FAS43138	DH2224/BC38	12	12	N	Y (19)	5,237,846	29 X
FAS43138	DH2228/BC43	9	9	N	Y (4)	5,159,317	28 X
FAS43138	DH1545/BC26	23	23	N	Y (21)	5,259,222	56 X
FAS43138	DH2235/BC20	7	7	N	Y (37)	5,384,176	21 X
FAS43138	DH2236/BC9	29	29	N	Y (6)	5,333,691	114 X

Nanopore flowcell ID's are given in the first column and isolate ID from the Department of Health UK study in the second column. Flye assembly contig number is given in parentheses and where this number is less than 10 the genome is usually closed.

3.3 Results

3.3.1 Genomic location of *bla*$_{CTX-M-15}$ in bacteraemia isolates

We investigated 95% (218/229) of the total number of *bla*$_{CTX-M-15}$ *E. coli* bacteraemia isolates collected in the ESBL UK study. These were collected from 5 regions of the UK and included 5 hospitals in East Anglia, 2 hospitals in the Northwest of England, Wales and Scotland and a single hospital in London (Day *et al* 2019). This represented 78% (229/292) of all ESBL *E. coli* bacteraemias from these sites. Other isolates encoded different ESBL genes such as *bla*$_{CTX-M-27}$ *bla*$_{CTX-M-14}$ and TEM ESBL variants. We located the genomic location of *bla*$_{CTX-M-15}$ by PFGE and in gel probing with *bla*$_{CTX-M-15}$ P^{32} labelled probes. A few isolates 4.6% (10/218) were negative for *bla*$_{CTX-M-15}$ and had lost the gene or plasmid carrying the gene either during growth or storage on Dorset egg slopes. We determined a chromosomal location for 56% (122/218) of isolates and a plasmid location for 47% (102/218) with 9% (20/218) carrying the *bla*$_{CTX-M-15}$ gene in both genomic locations (**Figures 3.3 and 3.4**). Chromosomally encoded isolates belonged to 15 different sequence types with 86 (70.5%) belonging to ST131 and 12 (8%) belonging to ST648. Other sequence types with a chromosomal *bla*$_{CTX-M-15}$ included ST38, 69, 70, 394, 405, 443, 501, 636, 746, 1193, 1196, 2851. Eighty-four per cent (102/122) of *E. coli* isolates with a chromosomally encoded *bla*$_{CTX-M15}$ gene did not have a plasmid version of the gene. Of the 9% of isolates with both chromosome and plasmid locations, 14 had a sole *bla*$_{CTX-M-15}$ plasmid. Whereas 4 and 2 isolates carried 2 and 3 *bla*$_{CTX-M-15}$ plasmids, respectively (**Figure 3.3**). Plasmid numbers ranged from 0-11 per isolate and varied in size from 25-480 kb with *bla*$_{CTX-M-15}$ plasmids ranging in size from 50- 400kb and 24% of *bla*$_{CTX-M-15}$ plasmids were *c.* 100kb (**figure 3.5**).

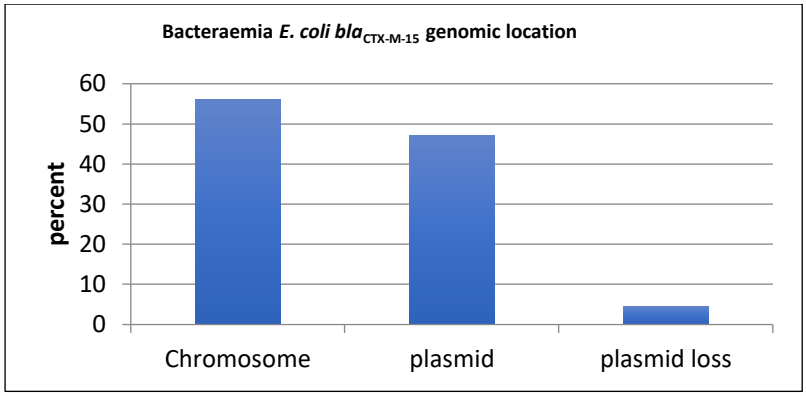


Figure 3.3. Genomic location of *bla*$_{CTX-M-15}$ genes in 218 human bacteraemia-derived *E. coli* isolates collected in 5 regions of the UK (August 2013-December 2014)

56% (122/218) of isolates were determined to be chromosomally encoded with 47% (102/218) isolates plasmid encoded. 4.6% (10/218) lost the *bla*$_{CTX-M-15}$ gene during growth (presumed plasmid loss).

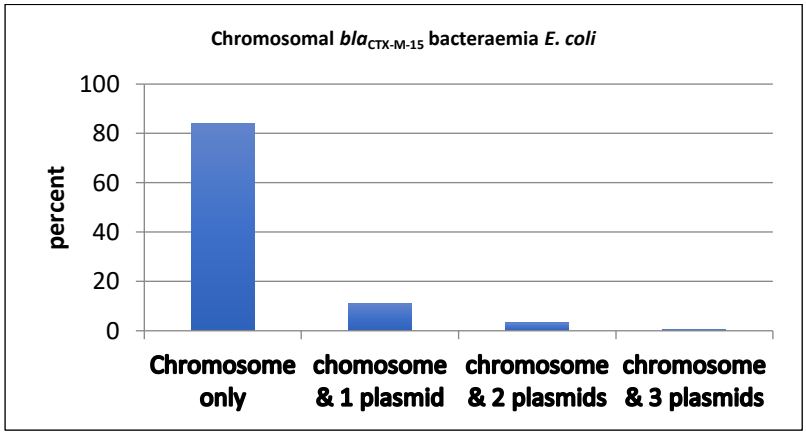


Figure 3.4. Breakdown of *E. coli* isolates with chromosomal encoded *bla*$_{CTX-M-15}$

84% (102/122) of isolates with a chromosomally encoded *bla*$_{CTX-M-15}$ gene did not contain a secondary plasmid version. 11% (14/122) had chromosomal and one plasmid carrying *bla*$_{CTX-M-15}$, 3% (4/122) and 0.6% with two and three *bla*$_{CTX-M-15}$ plasmids, respectively.

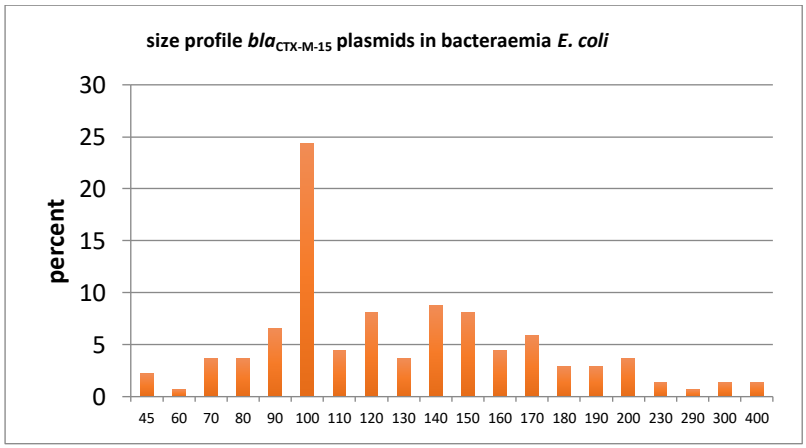


Figure 3.5. Size profile of *bla*$_{CTX-M-15}$ encoding plasmids in bacteraemia *E. coli* (collected in 5 regions of the UK August 2013- December 2015)

3.3.2 Genomic position of *bla*$_{CTX-M-15}$ in faecal *E. coli* isolates

We repeated our genomic analysis for the faecal *E. coli* isolates collected August 2013-December 2014 from the five regions in the UK. We investigated 97% (249/256) of all faecal CTX-M-15 encoding *E. coli* by PFGE and in-gel probing with a ^{32}P labelled *bla*$_{CTX-M-15}$ probe. Again, several isolates 9.2% (23/249) proved negative for the probe, likely due to plasmid loss during growth or transport/storage of the isolates. We found a slightly lower percentage of isolates with a chromosomal location for *bla*$_{CTX-M-15}$ than we found with the bacteraemia isolates collected at the same time. 49.4% (123/249) were chromosomally encoded and 48.9% (122/249) were plasmid-encoded with 4% (10/249) of isolates harbouring both plasmid and chromosomal copies (**Figure 3.6**). Chromosomal locations were found in 35 different sequence types with 50 (41%) belonging to ST131 and 8 (6.5%) belonging to ST38. Other sequence types included ST's 10, 46, 69, 70, 73, 155, 226, 349, 382, 405, 410, 443, 501, 602, 636, 648, 652, 746, 1193, 1312, 1325, 2551, 3052, 3075, 5667, 6288, 6291, 6294, 6302, 6309, and 6311. Among isolates having chromosomal and plasmid copies, 7 isolates had a single *bla*$_{CTX-M-15}$ plasmid and three had two *bla*$_{CTX-M-15}$ plasmids, all these belonged to ST131. Again, we found that many of the isolates with *bla*$_{CTX-M-15}$ on the chromosome (91.8%, 113/123) did not carry a plasmid version of the gene indicating subsequent plasmid loss after the capture of *bla*$_{CTX-M-15}$ by the chromosome **(Figures 3.6, 3.7, and 3.8**). Plasmids carrying *bla*$_{CTX-M-15}$ ranged in size from 40-350 kb with many around 100kb (**Figure 3.8)**. Among the 112 isolates carrying *bla*$_{CTX-M-15}$ plasmids, 88 carried just one *bla*$_{CTX-M-15}$ plasmid, 21 carried two, 2 carried three and one isolate carried 4 *bla*$_{CTX-M-15}$ plasmids. Analysis of short-read data indicated that most plasmids belonged to the *incF* incompatibility group.

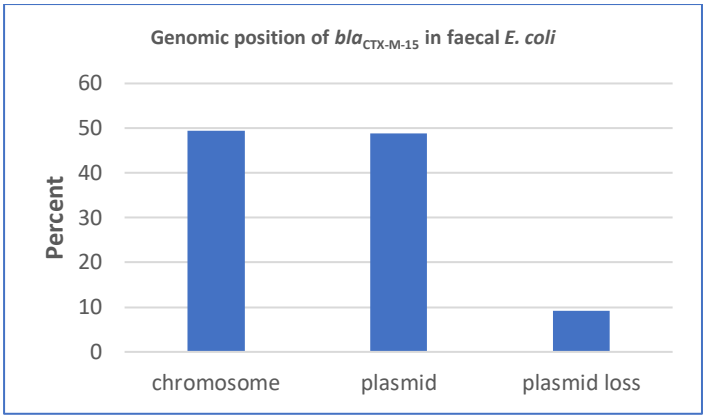


Figure 3.6. Genomic location of *bla*$_{CTX-M-15}$ in 249 human faecal *E. coli* isolates collected in 5 regions of the UK (August 2013-December 2014)

49.4% (123/249) were chromosomally encoded and 48.9% (122/249) were plasmid-encoded with 4% (10/249) of isolates harbouring both plasmid and chromosome copies.

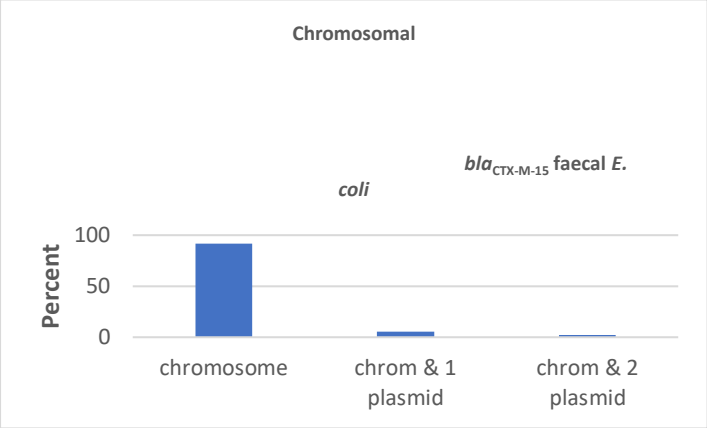


Figure 3.7. Breakdown of *E. coli* faecal isolates with chromosomal encoded *bla*$_{CTX-M-15}$

Many 91.8% (113/123) of the isolates with *bla*$_{CTX-M-15}$ on the chromosome did not carry a plasmid version of the gene indicating subsequent plasmid loss after the capture of *bla*$_{CTX-M-15}$ by the chromosome.

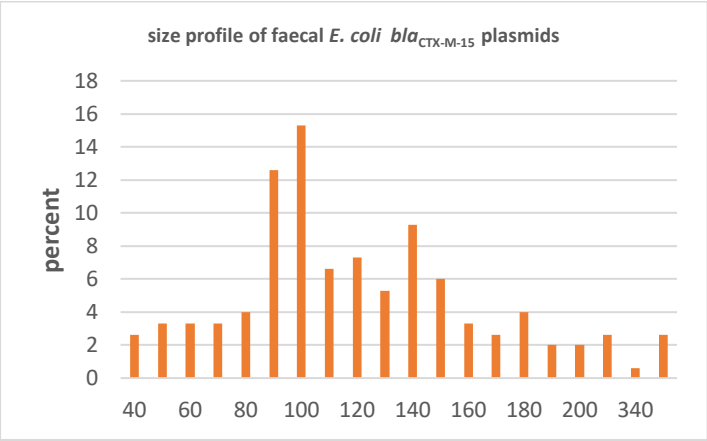


Figure 3.8. Size range of *bla*CTX-M-15 plasmids found in faecal *E. coli* collected in 5 regions of the UK (August 2013-December 2014)

3.3.3 Chromosomal insertion sites

The isolates harbouring chromosomal *bla*$_{\text{CTX-M-15}}$ genes were further interrogated to identify detail of the chromosomal insertion events. Since this was a large sample set, we chose a subset of 100 isolates (47%, 100/215). This included 48% (58/122) of the bacteraemia isolates and 34% (42/123) from faecal origin. Chromosomally located *bla*$_{\text{CTX-M-15}}$ genes were mapped in 10 different sequence types in bacteraemia isolates and from all regions of the UK. These included ST131 (28), ST648 (6), ST38 (2), ST405 (3), ST636 (2) and singletons of ST394, ST443, ST2851 and ST5667. Similarly, chromosomally encoded *bla*$_{\text{CTX-M-15}}$ genes were identified in faecal isolates from all regions of the UK. Our sample of 42 isolates included 17 different sequence types (20x ST131, 3x ST38, 2x ST443, 2x ST636, 2x ST648 and singletons of ST69, ST70, ST602, ST652, ST1193, ST2851, ST3075, ST5667, ST6289, ST6294 and ST6302.

3.3.4 Chromosomal insertion site analysis

Chromosomal insertion sites were located by local blast searches using Geneious software. This was to:

1. Determine the genetic vehicle(s) responsible for the movement event(s).
2. Determine the size of DNA mobilized.
3. To identify any other co-transposed genes.
4. To detect any “hot spots” of insertion onto the chromosome of *E. coli*.
5. To determine the nature of damage to the chromosomal genes and any detrimental or beneficial effects.

Among 100 isolates we identified 129 separate chromosomal insertions. Several isolates (21) had multiple chromosomal insertion sites ranging from 2 to 4 insertions

(**Table 3.3**). Isolates with multiple insertions were found in seven different ST from all regions of the UK. The *bla*$_{\text{CTX-M-15}}$ gene was mobilised mostly by IS*Ec*P1 in these isolates and insertions were of sizes 2796- 26,621 bp in length (**Table 3.3**). These multiple insertions were more common in bacteraemia isolates than in faecal isolates 14 vs 7.

Table 3.3. *E. coli* isolates with multiple chromosomal insertions of *bla*$_{CTX-M-15}$

Isolate ID	ST	Source (EA) E. Anglia, (L) London, (NW) N.West, (W) Wales, (S) Scotland.	Number	Size of inserts	Figure	Relevant mobile element
DH95	131	Bacteraemia (S)	4	5,289 bp 7,034 bp 20,673 bp 43,985 bp	**3.9.6** **3.11** **5.11.1**	IS*Ec*P1 IS26 ISEcp1 IS26
DH212	131	Bacteraemia (s)	2	2,976 bp 3,050 bp	**3.9.2** **3.9.2**	IS*Ec*P1 IS*Ec*P1
DH231	131	Bacteraemia (s)	2	2,981 bp 3,050 bp 7,9kb bp	**3.9.3** **3.9.3** **3.10**	ISEcP1 ISEcP1 complex
DH234	5667	Bacteraemia (s)	2	2,946 bp 3,055 bp	**3.9.2** **3.9.4**	IS*Ec*P1 IS*Ec*P1
DH246	131	Bacteraemia (NW)	2	4,755 bp 4,954 bp	**3.9.5** **3.9.6**	IS*Ec*P1 IS*Ec*P1
DH551	648	Faecal (NW)	3	2,794 bp 2,979 bp 4,808 bp	**3.9.1** **3.9.3** **3.9.5**	IS*Ec*P1 IS*Ec*P1 IS*Ec*P1
DH677	131	Faecal (W)	2	2,683 bp 10,534 bp	**3.9.1** **3.9.8**	IS*Ec*P1 IS*Ec*P1
DH726	131	Faecal (S)	2	28 kb 16,901 bp	**3.10**	Complex IS*Ec*P1
DH734	131	Faecal (S)	2	2,988 bp 4,780 bp	**3.9.3** **3.9.6**	IS*Ec*P1 IS*Ec*P1
DH804	131	Bacteraemia (EA)	2	23,530 bp	**3.9.10**	IS26
DH905	131	Faecal (L)	2	5,208 bp 12,958 bp	**3.9.6**	ISEcP IS26
DH964	6302	Faecal (L)	3	2,936 bp 3,053 bp 26,621 bp	**3.9.2** **3.9.4** **3.9.11**	IS*Ec*P1 IS*Ec*P1 IS26
DH1105	70	Faecal (EA)	2	4,783 bp 9,287 bp	**3.9.5** **3.9.7**	IS*Ec*P1 IS*Ec*P1
DH1535	131	Bacteraemia (W)	2	2,767 bp 6,037 bp	**3.9.1** **3.9.7**	IS*Ec*P1 IS*Ec*P1
DH1637	131	Bacteraemia (W)	4	2,796 bp 7,708 bp 10,550 bp 13,380 bp	**3.9.1** **3.9.7** **3.9.8** **3.9.9**	IS*Ec*P1 IS*Ec*P1 IS*Ec*P1 IS*Ec*P1
DH1687	2851	Bacteraemia (L)	3	2,796 bp 3,050 bp 4,344 bp	**3.9.1** **3.9.3** **3.9.5**	IS*Ec*P1 IS*Ec*P1 IS*Ec*P1
DH1688	648	Bacteraemia (L)	3	2,637 bp 8,965 bp 12,958 bp	**3.9.1** **3.9.7**	IS*Ec*P1 IS*Ec*P1 IS26
DH1696	648	Bacteraemia (NW)	2	7,374 bp 6,039 bp	**3.9.7** **3.9.7**	IS*Ec*P1 IS*Ec*P1
DH1705	131	Bacteraemia (W)	2	2,976 bp 24,550 bp	**3.9.3**	IS*Ec*P1 IS26
DH1707	394	Bacteraemia (W)	3	2,852 bp 2,855 bp 2,857 bp	**3.9.1** **3.9.1** **3.9.2**	IS*Ec*P1 IS*Ec*P1 IS*Ec*P1
DH2202	131	Bacteraemia (L)	2	1,795 bp 23,211 bp	**3.9.1**	IS*Ec*P1 IS26

Chromosomal insertions were mediated mostly by the transposable elements IS*Ec*P1 and IS*26*.

3.3.5 *ISEc*P-1 mediated chromosomal insertions

IS*Ec*P1 was responsible for most chromosomal insertion events (92) and these insertions ranged in size from 1,795 bp including just the IS*Ec*p1 element and a truncated section of the *bla*$_{\text{CTX-M-15}}$ gene to >26kb including multiple other antibiotic resistance genes and other transposable elements. Most of these insertions were accompanied by a target site duplication event evidencing the exact insertion size and that the insertion events were caused by transposition events. These direct repeats were found at the exact ends of the transposed section of DNA and were 5 bp in length and of high AT content, typical of what is known for this element. In most cases, where a direct repeat was not present, it was clear that recombinational events after the initial transposition event were responsible for this and these were often based on repetitive elements, notably IS26 found within the transposed section of DNA or sections of Tn2/3 and the same element found in another site within the bacterial chromosome. Detail of the insertions and co-transposed genes and direct repeats generated are given for *ISEc*P1 mediated insertions in (**Tables 3.4 & 3.5**) and in (**Figures 3.9.1- 3.9.11, and 3.10**) below. The insertions were found in both bacteraemia and faeces-derived isolates and from all regions of the UK. The tables and figures list the chromosomal insertions by insertion size. The first 44 insertions mobilise just the IS*EcP1* IS, the *bla*$_{\text{CTX-M-15}}$ resistance gene and a small section of a tryptophan synthetase gene and are below 4kb in size. Larger insertions include genes and gene fragments that the *ISEc*P1 transposon has collected and co-transposed by secondary transposition events from individual sites on the *E. coli* chromosome eg the *lysC* gene in DH1099 and DH1105 and the *dgoT* gene in

DH1687 (**Figure 3.9.5**). Mobilised regions up to 7,708bp include fragments of the Tn3 transposon and an IS26 element which likely originated from donating plasmids (**Figures 3.9.5-3.9.7**). Chromosomal insertions above 8kb more often include several other antibiotic resistance genes especially *bla*$_{OXA-1}$ and the bi-functional ***aac6'-Ib-cr*** gene that encodes both aminoglycoside and quinolone resistance and a truncated version of the chloramphenicol resistance gene *catA1* (**Figure3.9.7**) flanked by IS26 elements. Larger *ISEc*P1 chromosomal insertions have numerous antibiotic resistance genes co-transposed, often bracketed by IS26 elements including *bla*$_{TEM}$ genes and *aacC2* and *tmrB* genes **(Figure 3.9.8 & 3.9.9**). Several very large chromosomal *ISEc*P1 mediated insertions also include class 1 integrons that carry trimethoprim resistance genes such as *dfrA17* and aminoglycoside resistance gene cassettes especially *aadA5* and *aadB* and disinfectant resistance and sulphonamide resistance that is characteristic of the 3' conserved sequences of the class 1 integron (**Figure 3.9.10 & 3.10**). As the mobilised sections get larger and include more repetitive sequences such as several IS26 elements there is evidence of secondary homologous recombination events that have occurred since the original insertion event leading to loss of the direct repeat evidence of transposition and different pseudogenes formed at the ends of the insertions. **Figure 3.10** includes several very large chromosomal insertions that have a complex structure due to a combination of *ISEc*P1-mediated transposition, transposition of other elements and homologous recombination events. In the case of DH719. This includes the insertion of the transfer region of a plasmid and several other plasmid-encoded genes into a prophage region of the chromosome and may represent the formation of an *hfr* strain (**Figure 3.10**).

Table 3.4. IS*EcP1* mediated chromosomal insertions listed by size of insert

BACTEREMIA/FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TE M	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (Where multiple insertions are found in a single strain, the insertion number is given in bold)
DH2202	ST131	London	**3.9.1**	+	+		**1.** rlmA/DR/ISEcP1/DCTX-M-15/DR/rlmA	1,795 bp	ISEcP1	ND	*rlmA* gene. NB $bla_{CTX-M-15}$ is truncated
DH667	ST349	Wales	**3.9.1**	+	+		**1.** ISEcP1/CTXM-15/hypP	2,683 bp	ISEcP1	ND	Insert at end of contig therefore missing LHS sequence. Insertion is at p 7888/8082 of a large hypothetical protein encoding gene.
DH1688	ST648	London	**3.9.1**	+	+	+	**1.** pepT/DR/ISEcP1/CTX-M-15/DR/pepT	2,687 bp	ISEcp1	CTGGT	*pepT* at position 169/1227bp encoding a peptidase T protein.
DH1535	ST131	Wales	**3.9.1**	+	+		**1.** hlyC/DR/ISEcP1/CTX-M15/hypP/DR/dcuA	2,767bp	ISEcP1	TTATC	**1.** ISEcP-1 insertion into *dcuA* gene on RHE at position 833/1302 bp encoding a C4 dicarboxylate transporter but the rest of gene is not on the right-hand end suggesting recombination event since insertion. carrying only $bla_{CTX-M15}$ into the haemolysin gene.
DH551	ST648	N. West	**3.9.1**	+		+	**1.** rfaH/DR/ISECP1/CTX-M-15/DR/rfaH	2,794 bp	ISEcP1	GGTAT	**1.** ISEcp1 mediated into position 294/497bp of *rfa*H gene encoding a transcriptional/translational regulatory transformer protein.
DH520	ST131	N. West	**3.9.1**	+			bluF/DR/ISEcP-1/CTX-M-15/DR/bluF	2,758 bp	ISEcp1	TATAA	ISEcp1 mediated insertion into position 218/1212 of *bluf* encoding a blue-light responsive regulator BluR also called diguanylate phosphodiesterase.
DH1687	ST2851	London	**3.9.1**	+			**1.** ynfK/DR/ISEcP1/CTXM15/hypP/DR/ynfK	2,796 bp	ISEcP1	TTGCA	**1.** ISEcP1 mediated into position 445bp of *ynfK* gene encoding a putative dethiobiotin synthetase gene.
DH1637	ST131	Wales	**3.9.1**	+		+	**1.** artM/DR/ISEcP1/CTXM15/hypP/DR/artM	2,796 bp	ISEcP1	GAGTA	**1.** ISEcP1 mediated insertion into *artM* (arginine abc Transporter) at position 11/669 bp.
DH1707	ST394	Wales	**3.9.1**	+		+	**1.** pckA/DR/ISEcP1/CTXM15/hypP/DR/pckA	2,852 bp	ISEcP1	TTACT	**1.** ISEcP1 mediated insertion into position 210bp of pcKA gene encoding a phosphoenolpyruvate carboxykinase (ATP).
DH1707	ST394	Wales	**3.9.1**	+		+	**2.** cfa/DR/ISEcP1/CTX-M15/hypP/DR/cfa	2,855 bp	ISEcP1	TAATA	**2.** ISEcP1 mediated insertion into position 1051/1149 of *cfa* gene encoding cyclopropane fatty acyl phospholipid synthase M.
DH1707	ST394	Wales	**3.9.2**	+		+	**3.** pepT/DR/ISEcP1/CTXM15/hypP/tn3/DR/pepT	2,857 bp	ISEcP1	TTGTA	**3.** ISEcP1 mediated insertion at position 962/1226 bp of *pepT* gene encoding peptidase T.
DH782	ST131	E. Anglia	**3.9.2**	+	+	+	mppA/DR/ISEcP1/CTX-M-15/hypP/DR/mppA	2,887 bp	ISEcP1	TCAAA	*mppa* gene (p1614/1939 bp)
DH810	ST636	N. west	**3.9.2**	+		+	htrE/DR/ISEcP1/CTXM15/hypP/DR/htrE	2,916 bp	ISEcP1	TGATA	ISEcP1 mediated insertion into *htrA* gene encoding an outer membrane usher protein at position 1201/2589 bp.
DH1098	ST636	London	**3.9.2**	+			htrEDR/ISEcP1/CTXM15/hypP/DR/htrE	2,916 bp	ISEcP1	TGATA	ISEcP1 mediated insertion into *htrA* gene encoding an outer membrane usher protein at position 1201/2589 bp.
DH960	ST636	London	**3.9.2**	+			htrE/DR/ISEcP1/CTX-M-15/DR/htrE	2,916 bp	ISEcP1	TGATA	ISEcP1 mediated insertion into position 1201/2589bp of *htrE* gene encoding an outer membrane usher protein fim region
DH964	ST6302	London	**3.9.2**	+			**1.** nrfG/ISEcP-1/CTX-M-15/tn3/nrfG-cadC	2,936 bp	ISEcP1	ND	**1.** ISEcp1 mediated insertion into position 562/597bp of *nrfG* gene encoding a heme lyase.
DH234	ST5667	Scotland	**3.9.2**	+			**1.** rbsR/DR/ISEcP-1/CTX-M-15/DR/rbsR	2,946 bp	ISEcP1	AACTC	**1.** ISEcp1 mediated insertion into position 124/993bp of gene *rbsR* encoding a DNA binding transcriptional repressor.
DH722	ST652	Scotland	**3.9.2**	+		+	yjfO/DR/ISEcP-1/CTX-M-15/DR/yjfO	2,972 bp	ISEcP1	CCGTA	3' end of a lipoprotein encoding gene *yjfO* at position 325/350bp.
DH212	ST131	Scotland	**3.9.2**	+	+		**1.** uspC/DR/ISEcP1/CTX-M-15/DR/flhD	2,976 bp	ISEcP1	TAGCA	**1.** ISEcP1 mediated insertion is into an intergenic region between genes Universal stress protein *uspC* and flagellar transcriptional regulator *FlhD* .
DH456	ST131	N. west	**3.9.2**	+			uvrY/DR/ISEcP1/CTX-M-15/DR/uvrY	2,971 bp	ISEcP1	TAAAA	Miseq sequence are 118-110-21-90
DH1705	ST131	Wales	**3.9.3**	+	+	+	**1.** hypP/DR/ISEcP1/CTX-M-15/DR/hypP	2,976 bp	ISEcP1	AAATA	**1.** Insertion into hypP close to metG and molR fimbrial region.
DH97	ST405	Scotland	**3.9.3**	+			Cas3/DR/ISEcP1/CTX-M-15/tryp/DR/Cas3	2,976 bp	ISEcP1	AATAA	ISEcP1 mediated insertion at position 2244/2971bp of CRISPR-associated helicase. Similar sequences CP021732, CP021202
DH551	ST648	London	**3.9.3**	+		+	**2.** dacC /DR/ISECP1/CTX-M-15/DR/ gstB	2,979 bp	ISEcP1	TTTAC	**2.** ISEcp1 mediated found in intergenic region between *sstB* encoding a glutathione-s transferase and *dacC* encoding a serine-type D ala-D-ala carboxypeptidase(Penicillin binding protein).
DH941	ST3075	London	**3.9.3**	+		+	hypP/DR/ISEcP-1/CTX-M-15/DR/btuB	2,980 bp	ISEcP1	TTTTC	ISEcP1 mediated insertion between *hypP* and *btuB* gene encoding a Vitamin B12 transporter
DH231	ST131	Scotland	**3.9.3**	+		+	**1.** hypPDR/ISEcP1/CTX-M-15/DR/cirA	2,981 bp	ISEcP1	TTTAT	**1.** ISEcP1 mediated insertion into intergenic region In Phage related section of the chromosome. The insertion is directly adjacent and part of a larger chromosomal insertion (79,082bp) including a class 1 integron and several other resistance genes and the whole section is flanked by 2x cirA genes suggesting that the whole section has gone in by homologous recombination. Additional resistance genes are *sul*1/*aad*A2/*aac*C2/-*mph*A/*dfr*A12/*cat*A1/bla_{TEM-1b}. There are 4 copies of cirA in this strain that encode an outer membrane receptor involved in uptake of ferric dihyroxybenzoylserine also called TonB-dependent receptor.

Table 3.4 .IS*EcP1* mediated chromosomal insertions listed by size of insert (cont)

BACTEREMIA/FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TEM	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (where multiple insertions are found in a single strain, insertion number is given in bold)
DH734	ST131	London	**3.9.3**	+		+	**1. ISEcP-1/CTX-M-15/p fimbrial reg protein**	2,988 bp	IS*Ec*P1	TAAAA	IS*Ec*P1 mediated insertion into prophage region of the chromosome. Insertion is into region replete with IS and various phage genes.
DH826	ST5667	E. Anglia	**3.9.3**	+		+	**slrC/DR/ISEcP1/CTX-M-15/hypP/DR/slrC**	3,049 bp	IS*Ec*P1	TAAAC	gene slrC a PTS sorbitol transport subunit IIA gene at position 274/372bp. Insertion is identical to CP088790 and CP056282
DH212	ST131	Scotland	**3.9.3**	+	+		**2. phage/DR/ISEcP1/CTX-M-15/DR/phage**	3,050 bp	IS*Ec*P1	TCATA	**2.** IS*Ec*P1 mediated insertion into prophage on chromosome similar to phage in isolate CP088658. Insertion is into phage gene phage minor tail U family protein 12bp from end of the gene, 390bp from start codon.
DH1687	ST2851	London	**3.9.3**	+			**2. dgoT/DR/ISEcP1/CTX-M-15/tn3/DR/dgoT**	3, 050 bp	ISEcP1	TTATA	**2.** Insertion into position 233/1338 bp of dgoT aencoding a galactonate transporter
DH825	ST2851	E. Anglia	**3.9.3**	+			**acrF/DR/ISEcP1/CTX-M-15/DR/acrF**	3,050 bp	ISEcP1	TTAAT	Insertion into acrF gene
DH793	ST405	E. Anglia	**3.9.4**	+			**yacH/DR/ISEcP1/CTX-M-15/hypP/DR/yacH**	3,055 bp	IS*Ec*P1	ATTGA	P734/1758 yacH gene encoding.a putative mebrane protein.
DH897	ST131	E. Anglia	**3.9.4**	+			**hypP/DR/ISEcP-1/CTX-M-15/DR/hypP**	3,063 bp	ISEcP1	TCGAT	Position 12/393bp hypP gene upstream in Phage region, 753/1806 hypP downstream. 188kb distant in reference ST131 sequence.
DH964	ST6302	London	**3.9.4**	+			**2. ISEcP-1/CTX-M-15/tn3/acrA**	3,053 bp	IS*Ec*P1	ND	**2.** IS*Ec*p1 mediated insertion into position 974/1086 acrA gene encoding a MexE family multidrug efflux RND transporter periplasmic adaptor subunit. NB LHS is missing from the contig.
DH966	ST131	London	**3.9.4**	+			**rlmG/DR/ISEcP1/CTX-M-15/DR/rlmG**	3, 055 bp	ISEcP1	TAATA	ISEcP1 mediated insertion at position 1093/1142bp rlmG gene encoding a 23s RNA methylase
DH955	ST131	London	**3.9.4**	+			**rlmG/DR/ISEcP1/CTX-M-15/DR/rlmG**	3,055 bp	IS*Ec*P1	TATTA	ISEcP1 mediated insertion at position 1093/1142bp rlmG gene encoding a 23s RNA methylase
DH552	ST131	Wales	**3.9.4**	+			**rlmG/DR/ISEcP-1/CTX-M-15/tn3/DR/rlmG**	3,055 bp	IS*Ec*P1	TAATA	ISEcP1 mediated insertion into rlmG gene at position 1093/1142bp.
DH234	ST5667	Scotland	**3.9.4**	+			**2. slrD/DR/ISEcP-1/CTX-M-15/hypP/DR/slrD**	3,055 bp	IS*Ec*P1	TAAAC	**2.** IISEcp1-mediated insertion into p 267/1154bp of slrD gene encoding a sorbitol-6-phosphate dehydrojenase.
DH458	ST131	Northwest	**3.9.4**	+		+	**mppa/DR/ISEcP1/CTX-M-15/DR/mppa**	3,217 bp	IS*Ec*P1	TCAAA	ISEcP1 mediated insertion into gene oligopeptide ABC transporter substrate-binding protein OppA also called murein tripeptide ABC transporter OppBCDFMppA - periplasmic binding protein 330bp from end 0f gene. Identical insertion found in strain CP023853. Insertion is close to the universal stress protein uspC.
DH2205	ST131	London	**3.9.4**	+		+	**mppa/DR/ISEcP1/CTX-M-15/hypP/DR/mppa**	3,217 bp	IS*Ec*p1	TCAAA	ISEcP1 mediated into position 1284/1614 bp of mppa gene encoding an peptide ABC transporter substrate binding protein near the universal stress protein uspA. NB similar sequence carries similar insertion CP023853 also identical to DH458 above.
DH834	ST443	E. Anglia	**3.9.4**	+			**iraM/DR/ISEcP-1/CTX-M-15/tn3/DR/iraM**	3,279 bp	ISEcP1	ATTCA	ISEcp-1 mediated insertion of 3279bp into position 272/2336 Of iraM gene encoding an anti-adapter protein IraM. Again insertion is into a prophage region.
DH856	ST443	E. Anglia	**3.9.5**	+			**iraM/DR/ISEcP-1/CTX-M-15/tn3/DR/iraM**	3,279 bp	ISEcP1	TGAAT	Inserted into position 247/311 bp of anti adapter protein encoding gene that is part of a large Phage region.
DH78	ST443	Wales	**3.9.5**	+			**iraM/DR/ISEcP-1/CTX-M-15/tn3/DR/iraM**	3,279 bp	ISEcP1	TGAAT	Inserted into position 247/311 bp of anti adapter protein encoding gene that is part of a large Phage region.
DH857	ST602	E. Anglia	**3.9.5**	+			**yehB/DR/ISEcP-1/CTX-M-15/tn3/DR/yehB**	3,506 bp	ISEcP1	TGTCC	ISEcp1 mediated insertion into yehB gene encoding an outer membrane usher protein in a fimbrial operon. Insert is at position 1350/2024bp
DH84	ST1196	N. West	**3.9.5**	+			**hypP/DR/ISEcP1/CTX-M-15/Tn3/DR/DNAbind**	3,747 bp	ISEcP1	ATAAT	Intergenic insertion between hypP and a DNA binding protein encoding gene.
DH1687	ST2851	London	**3.9.5**	+			**3. dcuS/DR/ISEcP1/CTX-M-15/dgoT/DR/dcuS**	4,344 bp	ISEcP1	GAACA	**3.** Insertion into position 1522/1527 of dcuS encoding a sensor histidine kinase protein of a 2-component regulator.
DH246	ST131	N. West	**3.9.5**	+	-	+	**1. GTPase/ISEcP1/CTX-M-15/Tn3/lysC2/hmuT**	4,755 bp	ISEcP1	TGTTA	**1.** Intergenic region between a prophage hypothetical gene (CP4-57) and a hmuT gene encoding a hemin binding periplasmic protein. NB inserts are related as same section of DNA mobilised and insertion site on one side is identical see blast hits in AP sequence
DH1099	ST131	Scotland	**3.9.5**	+			**GTPase/DR/ISECp1/CTX-M15/hypP/tn3/lysC/DR/Phage**	4,783 bp	ISEcP1	TGTTA	ISEcp-1 mediated within prophage on the chromosome. Insertion size 4763bp between 50s ribosome binding gtpase And other phage CP4-57 genes -intergenic insertion.Mobilised section includes ctx-m15/tn3/aspartate kinase gene lysC.

Table 3.4. IS*EcP1* mediated chromosomal insertions listed by size of insert (cont)

BACTEREMIA/FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TEM	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (where multiple insertions are found in a single strain, insertion number is given in bold)
DH1105	ST70	E. Anglia	**3.9.5**	+			GTPase/DR/ISECp1/CTX-M15/hypP/tn3/lysC/DR/Phage	4,783 bp	ISEcP1	TGTTA	ISEcp-1 mediated within prophage on the chromosome. Insertion size 4763bp between 50s ribosome binding gtpase And other phage CP4-57 genes -intergenic insertion.Mobilised section includes ctx-m15/tn3/aspartate kinase gene lysC.
DH60	ST405	Wales	**3.9.5**	+			nudL/DR/ISEcp1/CTX-M-15/Tn3/DR/nudL	4,804 bp	ISEcp1	TATAA	Position 27/187 nucleoside triphosphase gene nudL
DH551	ST648	N. West	**3.9.5**	+		+	**3.** ypdC/DR/ISEcP1/CTX-M-15/tn3/DR/ypdC	4,808 bp	ISEcP1	TAAAA	**3. I**nsertion intop position623/864 of gene ypdC encoding a putative ARAC type regulatory protein.
DH734	ST131	Scotland	**3.9.6**	+		+	**1.**	4,780 bp	ISEcP1		Inserted into transposon gene
DH540	ST6289	N. West	**3.9.6**	+		+	hypP/DR/ISEcP1/CTX-M-15/tn3/DR/hypP	4,810 bp	ISEcP1	TGATA	Insertion into phage hypP at position 385/2175.
DH951	ST6288	London	**3.9.6**	+	+	+	hypP prophage/DR/ISEcP-1/CTX-M-15/hypP/tn3/DR/ hypP prophage	4,812 bp	ISEcP1	TGATA	ISEcp1 mediated into position 384/2517bp of hypothetical protein gene within a prophage region of the Chromosome. Very similar insertion into genome CP099754
DH246	ST131	N. West	**3.9.6**	+		+	**2.** cirA/ISEcP1/CTX-M-15/Tn3/lysC2/yfjI	4,954 bp	ISEcP1	TGATA	**2.** Intergenic region between cirA gene encoding a hemin receptor and a bacteriophage gene yfjI (CP4-57)
DH536	ST6294	N. West	**3.9.6**	+			ypjF//ISEcP1/CTX-M-15/Tn3/Tn3?	5,750 bp	ISEcP1	ND	Position 343/369 gene ypjF encoding CP4-57 phage toxin of toxin /antitoxin system.
DH905	ST131	London	**3.9.6**	+	+		**1.** ydhA/DR/ISEcP1/CTX-M-15/tn3/DR/ydhA	5,208 bp	ISEcp1	TATGA	**1.** Position 4/324 of gene ydhA similar to DH677 and DH1535 ydh encodes a lysozyme inhibitor
DH95	ST131	Scotland	**3.9.6**	+	-		**1.** motA/ISEcP-1/CTX-M-15/tn3/hypP/motA	5,289bp	ISEcP1	GTATA	ISEcp1 mediated insertion into motA gene at position 229/888bp encoding a flagellar motor stator protein.
DH228	ST131	Scotland	**3.9.6**	+	-		motA/ISEcP-1/CTX-M-15/tn3/hypP/motA	5,923 bp	ISEcP1	GTATA	ISEcp1 mediated insertion into motA gene at position 229/888bp encoding a flagellar motor stator protein.
DH240	ST636	N. West		+	-	+	hypP/ISEcp1/CTX-M-15/?	3,700 bp	ISEcP1	ND	ISEcp-1 mediated insertion into a hypothetical Protein. RHS truncated by contig.
DH356	ST131	N. West	**3.9.7**	+			creC/ISEcP1/CTX-M-15/tn3/IS26/flagella genes	6,035 bp	ISEcP1	ND	ISEcp-1 mediated insertion into creC (two-component system sensor histidine kinase CreCat) at position 797/1425. Insertion includes CTX-M15 and section of tn3 and IS26. No direct repeat sequence found at junction with adjacent DNA therefore probably changed by IS26 based recombination events subsequent to insertion. NB the flanking sequences are far apart in the highest hit seq from blast searchs ee subfolder for detail.
DH1535	ST131	N. West	**3.9.7**	+		+	**2.** ydhA/ISEcP1/CTXM15/hypP/tn3/IS26/dhaR	6,037 bp	ISEcP1/ IS26	ND	**2.** ISEcP1 mediated with subsequent homologous recombination based on IS26 at RHE ISEcp-1 at LHE inserted at p3/324bp gene ydhA encoding a lysozyme inhibitor. RHE is inserted at position 690/1842 of transcriptional regulator gene ofsigma-54-dependant Fis family Genes on AP026518 are far apart (323,304-4883229).
DH1696	ST502	N. West	**3.9.7**	+		+	**2.** RecQ/rhlE/DR/ISEcP1/CTXM15/hypP/tn3/IS26/ hypPx2/TEM/DR/RecQ/rhlE	7,374 bp	ISEcP1	ATTAT	**2.** ISEcP1 mediated (region1) insertion into 1515/2097bp of rhlE gene encoding an ATP dependant helicase. Similar/same insertions also found in strains DH1681 and DH1647.
DH1090	ST131	Wales	**3.9.7**	+	+	+	1.ggt/DR/ISEcP1/CTXM15/hypP/tn3/insD/insC/ DR/pentapeptide	7,225 bp	ISEcP1	TGGAA	**1.** Inserted into phage/plasmid region on the chromosome between ggt and pentapeptide repeat genes in region with numerous insertion sequences.
DH1696	ST502	N. West	**3.9.7**	+		+	**2.** hypP/DR/ISEcP1/CTXM15/hypP/tn3/IS26/ DR/res	6,039 bp	ISEcP1	ND	**2.** ISEcP1 mediated insertion into hypothetical gene position 321/336on LHE and type 111 restriction enzyme at RHE at position 1454/4031. Good ST648 genome for comparisons CP008697.
DH1637	ST131	Wales	**3.9.7**	+		+	**2.** ynfD/DR/ISEcP1/CTXM15/hypP/tn3/IS481/ DR/ynfE	7,708 bp	ISEcP1	TCATA	**2.** insertion into intergenic region between ynfD and ynfE genes encoding a membrane protein and dimethyl sulfoxide reductase subunit A.
DH1688	ST648	London	**3.9.7**	+	+	+	**2.** Phageint/DR/ISEcP1/CTX-M-15/tn3/IS26/cat/OXA/aac/IS26/DR/Phageint	8, 965 bp	ISEcP1	TATAC	Insertion into p4/1206 bp of phage integrase gene.
DH1105	ST70	E. Anglia	**3.9.7**	+			**2.** DdlA/DR/ISEcP1/CTXM15/hypPx3/tn3/hypPx4/ DR/iraP	9, 287 bp	ISEcP1	TAATT	**2.** ISEcp-1 mediated insertion identical to section from several plasmids Also similar to section on chromosome of isolates from Bangladesh but not annotated. Insertion into intergenic region between ddlA encoding D-alanine D- alanine ligase and iraP encoding and anti-adapter protein.
DH1547	ST648	N. West	**3.9.7**	+	+	+	RecQ/rhlE/DR/ISEcP1/CTX-M15/hypP/tn3/IS26/catB3/OXA-1/aac6'-Ib-cr/IS26/TEM12/DR/RecQ/rhlE	10,136 bp	ISEcP1	TAATA	ISEcP1 mediated insertion into rhlE gene encoding a recQ family ATP Dependant helicase at position 1516/2097 bp of the gene.

Table 3.4 IS*EcP1* mediated chromosomal insertions listed by size of insert (cont)

BACTEREMIA/FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TEM	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (where multiple insertions are found in a single strain, insertion number is given in bold)
DH1647	ST131	Wales	**3.9.7**	+	+	+	RecQ/rhlE/DR/ISEcP1/CTXM15/hypP/tn3/IS26/cat/OXA1/aac6'-Ib-cr/IS26/TEM/DR/RecQ/rhlE	10,136 bp	ISEcP1	TAATA	ISECP1 mediated large insertion 10,136bp into rhlE gene encoding an ATP-dependant RNA helicase at position 1511/ 2097 bp.
DH677	ST131	Wales	**3.9.8**	+		+	**1.** ydhA/DR/ISEcP1/CTX-M-15/Tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/CAT/oxa-1/DR/ydhA	10,543 bp	ISEcP1	TATGA	**1.** Position 4/ 324 ydhA gene/mliC gene. NB OXA gene is truncated in the middle of the gene by transposition event
DH1637	ST131	Wales	**3.9.8**	+		+	**3.** ydhA/DR/ISEcP1/CTX-M-15/Tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/CAT/oxa-1/DR/ydhA	10,550 bp	ISEcP1	TATGA	**3.** Position 4/ 324 ydhA gene/mliC gene. NB OXA gene is truncated in the middle of the gene by transposition event-Exact insertion as above.
DH902	ST69	London	**3.9.8**	+	+	+	rutD/SEcP1/CTX-M-15/Tn3/IS26/ISKpn11/tmrB/aacC2/IS26/CAT/oxa-1/aac6'-Ib-cr/rutD	11,150 bp	ISEcP1	ND	rutD (p391/ 813)
DH2252	ST38	Wales	**3.9.8**	+			nimT/DR/ISEcP1/CTX-M-15/Tn3/IS26/hypP / ISKpn11/hypP/DR/nimT	11,390 bp	ISEcP1	ND	Phage gene nimT
DH1134	ST1193	London	**3.9.8**	+		+	yecD/ISEcP-1/CTX-M-15/tn3/IS26/aac6'-Ib-cr/ OXA-1/cat/IS26/hypP/hypP/tnrB/ISkpn11/ IS26/ /DR/yecD	11,740 bp	ISEcP1	TGTTA	ISEcp1 mediated insertion of 11,740bp into yecD gene encoding a putative hydrolase at position 378/566 bp.
DH2220	ST131	London	**3.9.8**	+	+	+	Dam1/DR/ISEcP1/CTXM15/hypP/tn3/IS26/ISKPn11/tmrB/aacC2/IS26/catB3 /OXA-1/ aac6'-Ib-cr IS26/DR/hypP	12,049 bp	ISEcP1	ATTAA	ISECP1 mediated insertion between dam1 restriction endonuclease and a hypothetical protein. The final section of an IS26 elementas been transposed but the terminus inverted repeat sequence has been truncated.
DH935	ST131	London	**3.9.8**	+	+	+	**1.** FAD-NAD/ISEcP1/CTX-M-15/Tn3/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/FAD-NAD	12,701 bp	ISEcp1	GCAAA	**1.** Insertion is into a FAD/NAD (p) domain containing protein gene at position 698/1605bp.
DH1637	ST131	Wales	**3.9.9**	+	+		**4.** IldD/DR/SEcP-1/CTX-M15/tn3/IS481/insF/hypP/ ISEc23/betI/hypP/yjcD	13,380 bp	ISEcP1	TTATA	**4.** Insertion between genes IldD and yjcD encoding L-lactate dehydrojenase and CPA1 transporter respectively.
DH2227	ST38	London	**3.9.9**	+	+		Ail/lom/DR/ISEcP1/CTX-M15/tn3/tn3/IS26/ hypPx2/tmrB/aacC2/IS26/aac6'-1b-cr/ OXA/cat/ IS26/DR/Ail/Lom	13,557 bp	ISEcP1	ND	Insertion into Ail/lom phage gene
DH785	ST648	E. Anglia	**3.9.9**	+	+	+	RecQ/rhlE/ISEcP1/CTX-M-15/tn3/IS26/aac6'-iB-cr/OXA-1/cat/IS26/hypPx4/IS150/IS26/ TEM/ recQ/rhlE	13,564 bp	ISEcP1	TAATA	recQ (p1517 of 2097bp) as below
DH1681	ST648	London	**3.9.9**	+	+	+	RecQ/rhlE/DR/ISEcP1/CTXM15/hypP/tn3/IS26/ISKpn11/tmrB/aacC2/cat/OXA1/aac6'-Ib-cr/IS26 /TEM/DR/recQ/rhlE	13,564 bp	ISEcP1	TAATA	ISECP1 mediated large insertion 10,136bp into rhlE gene encoding an ATP-dependant RNA helicase at position 1511/ 2097 bp. ISEcP1 mediated
DH104	ST648	E. Anglia	**3.9.9**	+	+	+	RecQ/rhlE/DR/ISEcP1/CTX-M-15/Tn3/IS26/ ISKpn11/ tmrB/aacC2/IS26/cat/oxa-1/aac6'-Ib-cr/ IS26/TEM/DR/RecQ/rhlE	13,564 bp	ISEcP1	TAATA	rhlE gene (p1517/2097bp)
DH107	ST38	E. Anglia	**3.9.9**	+	+	+	pepT/DR/ISEcP1/CTXM15/tn3/IS26/aac6'-ib-cr/OXA-1/catB3/IS26/IS26/pinQ/ tn3/chpB/ pemL /DR/pepT	14,093 bp	ISEcP1	GTAAT	pepT (p169/1227 bp)
DH763	ST131	Scotland	**3.9.10**	+			Sat/IS26/hypP/class1int/dhfr/ant1/qacE/sul1/hypP/IS110/srpC/hypP/IS6100/tap/hypP/IS26/ISEcP1/CTX-M-15/prophage	15,830 bp	ISEcP1	ND	CTX-M-15 was mobilised by ISEcp1 originally but the LH end has been truncated by IS26 since the original insertion event. Insertion at the LHE is into position 2884/3888 of toxin gene sat which is within a prophage the right hand cannot be determined exactly but also is within a prophage region.
DH923	ST131	N. West	**3.9.10**	+	+		Phagerep/DR/ISEcP1/CTX5/hypP/tn3/tn3/IS26/hypP/OXA1/hypPx3/tn3/tetR/yajR2/tetA/hypPx2/tn3/pinR /tn3/chpB/pemI/hypP/DR/Phage rep	19,160 bp	ISEcP1	TTTATA	ISEcP1-mediated insertion into position 422/786 bp of a phage replication protein gene. Genes most similar to LR134157.
DH956	ST131	London	**3.9.10**	+	+	+	nanC/ISEcP1/CTX-M-15/Tn3/IS26/ISKpn11/ tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/IS26 tn3/hypP/tetA/yajR/tetR/tn3/IS26/fecI	19,209 bp	ISEcP1	ND	ISEcp1 mediated insertion into position 10/1008 nanS gene. nanM upstream encodes a N-acetylneuraminate epimerase nanS is not found at the other end but rather a fec operon so that inserted section is found between fimbrial and fec operons.

Table 3.4. IS*EcP1* mediated chromosomal insertions listed by size of insert (cont)

BACTEREMIA/FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TE M	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (where multiple insertions are found in a single strain, insertion number is given in bold)
DH771	ST131	E. Anglia	**3.9.10**	+	+	+	IS26/tsv2/hypP/IS26/class1-int/dfrA17/aadA5/ qac/sul1/hypP/srpC/IS6100/hypPx2/ tap2/hypP/ IS26/pinE/tn3/IS1380/ISEcP-1/CTX-M-15	20,383bp	ISEcP1	ND	Exact insertion site cannot be determined because of truncation Similar to several others in genbank *Escherichia coli* strain ESBL179 chromosome Sequence ID: CP041559.1 *Escherichia coli* strain ESBL193 chromosome Sequence ID: CP041556.1 *Escherichia coli* strain JJ1897 plasmid pJJ1897_1, complete sequence Sequence ID: CP013836.1
DH927	ST38	London	**3.9.10**	+	+	+	Phage rep/DR/ISEcp1/CTX-M-15/tn3/tn3/IS26 /cat/OXA-1/aac6'-Ib-cr/IS26/tn3/tetR/yajR /tetA/ hypPx2/tn3/pinR/chpB/pemL/hypPx2/DR/Phage rep	20,619 bp	ISEcP1	TTATA	insertion into putative phage rep protein encoding gene at position 422/786bp.
DH754	ST131	Scotland	**3.9.10**	+	+	-	yedA/ISEcP-1/CTX-M-15/hypP/tn3/IS26/ hypPx4/ IS26/folP/neo/hypP/tetR/hypP/hypP /IS26/IS4 /hypP/IS26/TEM/hypP	c. 22kb	ISEcP1	ND	Likely ISEcp1 mediated insertion of CTX-M-15 into large ISEcp1 insert is into yedA 689/921 gene encoding a putative inner membrane transporter but other end of the gene is missing suggesting change since insertion.
DH804	ST131	E. Anglia	**3.9.10**	+	+	+	1. linoA/DR/ISEcP-1/CTX-M-15/tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/ IS26/ hypPx2/tetR/hypPx2/tetA/hypP/tn3/pinR/tn3/ chpB/pem/hypP/DR/linoA	23,530 bp	ISEcP1	ATAAC	ISEcp1 mediated insertion of 23,530bp into a linoA gene encoding an linoleoyl-CoA desaturase position 102/1022bp.
DH2206	ST131	London	**3.9.10**	+	+	+	linoA/DR/ISEcP-1/CTX-M-15/tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr /IS26/ hypPx2/tetR/hypPx2/tetA/hypP/tn3/pinR /tn3/chpB/pem/hypP/DR/linoA	23,530 bp	ISEcP1	ATAAC	ISEcp1 mediated insertion of 23,530bp into a linoA gene encoding an linoleoyl-CoA desaturase position 102/1022bp.
DH2229	ST131	London	**3.9.10**	+	+	+	linoA/DR/ISEcP-1/CTX-M-15/tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/cat/OXA-1/ aac6'-Ib-cr/IS26 /hypPx2/tetR/hypPx2/tetA/hypP/tn3/pinR /tn3/chpB/pem/hypP/DR/linoA	23,530 bp	ISEcP1	ATAAC	ISEcp1 mediated insertion of 23,530bp into a linoA gene encoding an linoleoyl-CoA desaturase position 102/1022bp.
DH2230	ST131	London	**3.9.11**	+	+	+	linoA/DR/ISEcP-1/CTX-M-15/tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/cat/OXA-1/ aac6'-Ib-cr/IS26 /hypPx2/tetR/hypPx2/tetA/hypP/tn3/pinR /tn3/chpB/pem/hypP/DR/linoA	23,530 bp	ISEcP1	ATAAC	ISEcp1 mediated insertion of 23,530bp into a linoA gene encoding an linoleoyl-CoA desaturase position 102/1022bp.
DH2238	ST131	London	**3.9.11**	+	+	+	linoA/DR/ISEcP-1/CTX-M-15/tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/cat/OXA-1/ aac6'-Ib-cr/IS26 /hypPx2/tetR/hypPx2/tetA/hypP/tn3/pinR /tn3/chpB/pem/hypP/DR/linoA	23,530 bp	ISEcP1	ATAAC	ISEcp1 mediated insertion of 23,530bp into a linoA gene encoding an linoleoyl-CoA desaturase position 102/1022bp.
DH2240	ST131	London	**3.9.11**	+	+	+	linoA/DR/ISEcP-1/CTX-M-15/tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/cat/OXA-1/ aac6'-Ib-cr/IS26 /hypPx2/tetR/hypPx2/tetA/hypP/tn3/pinR /tn3/chpB/pem/hypP/DR/linoA	23,530 bp	ISEcP1	ATAAC	ISEcp1 mediated insertion of 23,530bp into a linoA gene encoding an linoleoyl-CoA desaturase position 102/1022bp.
DH2218*	ST131	London	**3.9.11**	+	+	+	hypP/DR/ISEcP1/CTXM15/hypP/tn3/IS26/ISKPn11/tmrB/aacC2/IS26/catB3/IS26/aac6'-Ib-cr/OXA-1/IS26/tn3/tetR/yajR/tetA/hypP/tn3/pinR/tn3/chpB/peml/hypP/DR/hypP	24,360 bp	ISEcp1	ATAAC	ISEcp1 mediated into a hypP gene see insert in AP026518 which is similar to 50 and DH2206.
DH964	ST6302	London	**3.9.11**	+			rfbB/DR/ISEcP1/CTXM15/tn3/hypP/pemL/chbP/tn3/pinR/Tn3/hypP/tetA/yjha/tetR/tn3/IS26 /hypP/ vapC/hypPx4/vapC/hypPx3ISEc23/DR/rfbB	26,621	ISEcP1	GAATA	ISEcP1 mediated into position 9/900 bp of rfbB gene encoding a NAD(P) dependant oxidoreductase gene.

Table 3.5. Large chromosomal insertions of unclear origin

BACTEREMIA/FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TE M	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (where multiple insertions are found in a single strain, insertion number is given in bold)
DH726	ST131	Scotland	**3.10**	+			IS26/agP1/hypP/resA/hypP/ybbA1/hypPx3/tpd/hypPx2/nqrC/IS2/hypP/xerD/xerC/ant1/qacE/sul1/hypP/IS110/srpC/hypP/IS6100/tap/hypP/IS26/ISEcP1 /CTX-M-15	c. 28kb	ISEcP1/ IS26	ND	CTX-M-15 found in a resistance island Closely associated with ISEcp-1 so it probably got onto the chromosome via this element but subsequently the LH terminus (76bp) ofISEcp-1 has been truncated by IS26. The region includes Several plasmid genes and is similar to CP054215. Upstream is an phage integrase adjacent to a tRNA therefore this section is likely of phage origin and includes both antibiotic resistance and virulence genes- perhaps an ICE though no obvious mob genes are present.
DH231	ST131	Scotland	**3.10**	+		+	cirA/hypPx3/IS100/IS21/ISCro/hypPx3/papB/cafM/ effC/hypP/rutG/rutF/rutE/rutD/rutC/rutB/rutA/ hypP/putP/eteU/hypP/eteB/phoH/paaB/paaC/ paaB/paaA/ycdT/hypP/phage-int/yfiL/hypPx3/insA/catA1/ hypP/Tn3/IS26/class1-int/dfrA12/aadA2/qac/sul/ hypP/srpc/IS6100/hypPx2/tap/mphA/IS26/aacC2/tmrB/hypPx2/IScfr/blaTEM/hypPx2/ISEc45/ISEcP1/CTX-M-15/cirA	>79kb	Homologous recombination	ND	ISEcP1 mediated insertion into hypP encoding gene In Phage related section of the chromosome. Homologous recombination. The ISECP-1 mediated insertion is directly adjacent and part of a larger chromosomal insertion (79,082bp) including a class 1 integron and several other resistance genes and the whole section is flanked by 2x cirA genes suggesting that the whole section has gone in by homologous recombination. Additional resistance genes are sul1/aadA2/aacC2/-mphA/dfrA12/catA1/TEM-1b. There are 4 copies of cirA in this strain that encode an outer membrane receptor involved in uptake of ferric dihyroxybenzoylserine also called TonB-dependent receptor.
DH719	ST131	Scotland	**3.10**	+	-	+	Phage CP4-57 genes/CTX-M-15/ISEcp-1/IS26/hypP/tap/hypP/IS6100/srpC/sul/qacE/ant/ dhfr/class1int/IS26/tn3/tetR/yajR/tetA/hypPx2/tn3/pinR/tn3/hypP/chpB/pemL/hypPx3/ISEc23/hypP/finO/Plasmid transfer reg/Phage CP4-57 related genes		ISEcp1/IS26/hfr-like	ND	ISEcp1 mediated insertion of CTX-M-15 but this whole section Is actually an entire plasmid that has inserted onto the Chromosome ie 85 is a hfr strain. The plasmid insertion also Appears to be within a prophage located on the chromosome Similar to CP4-57. The CTX-M-15 is part of a larger resistance Island.

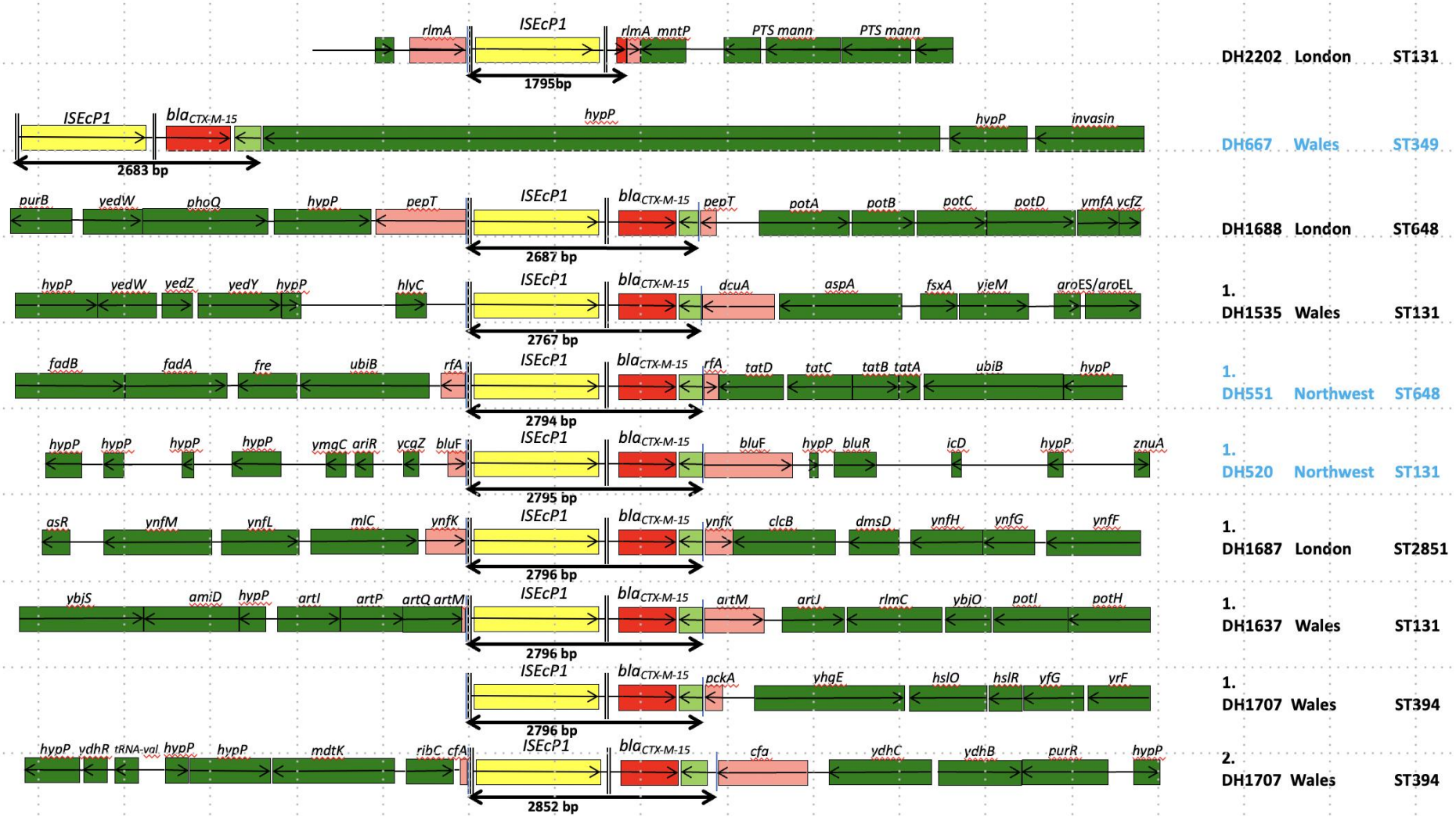


Figure 3.9.1. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

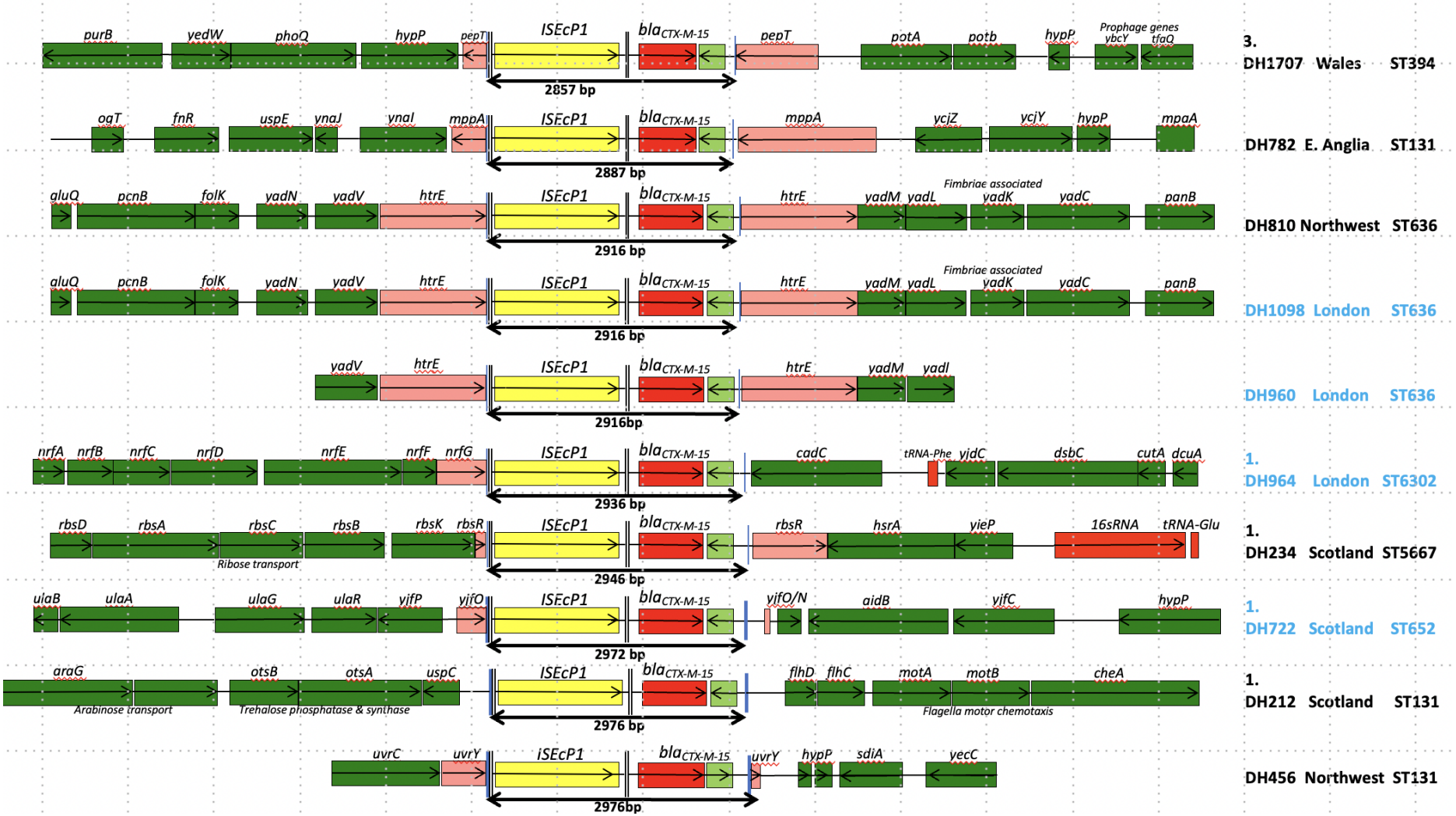


Figure 3.9.2. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

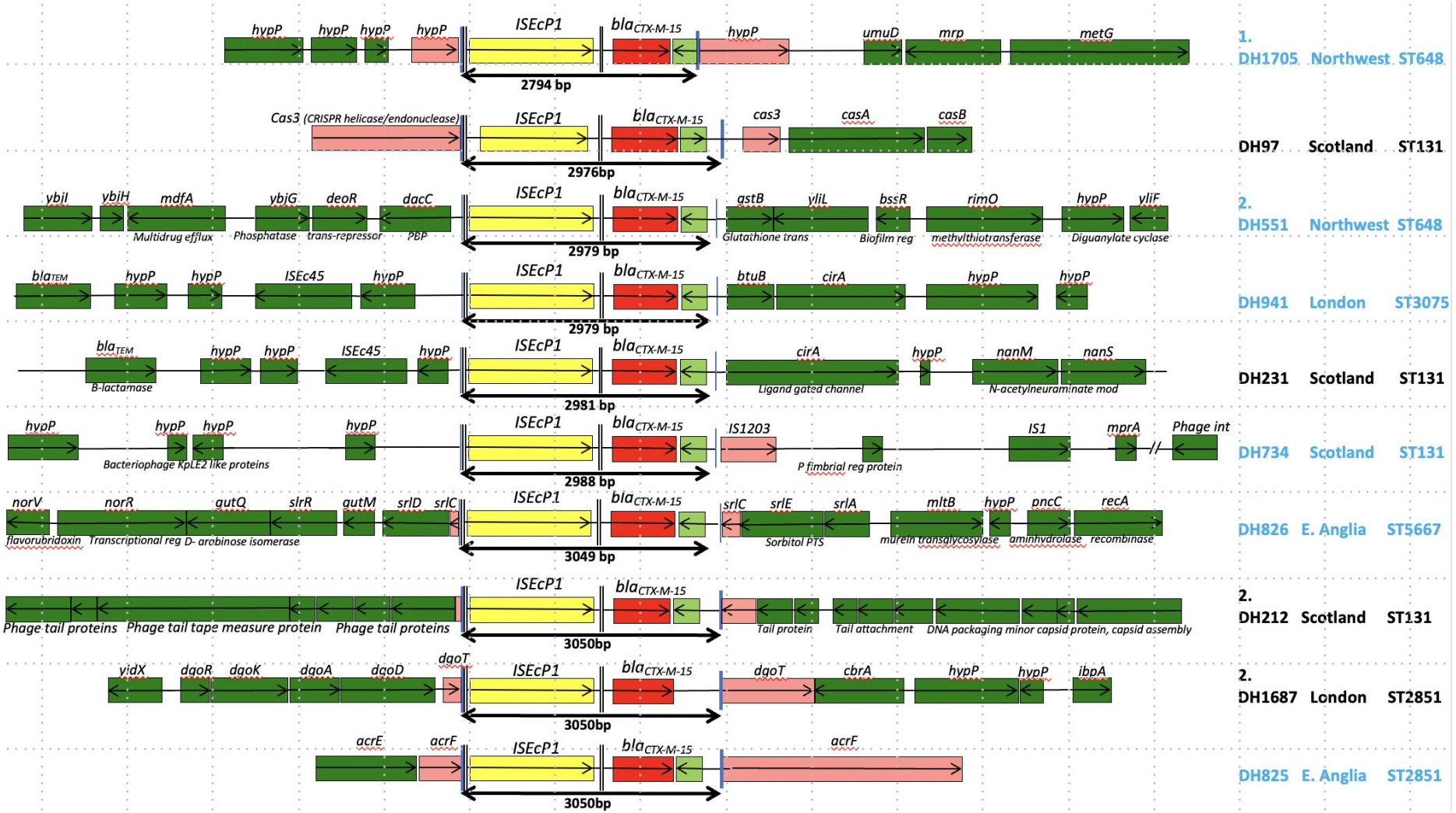


Figure 3.9.3. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

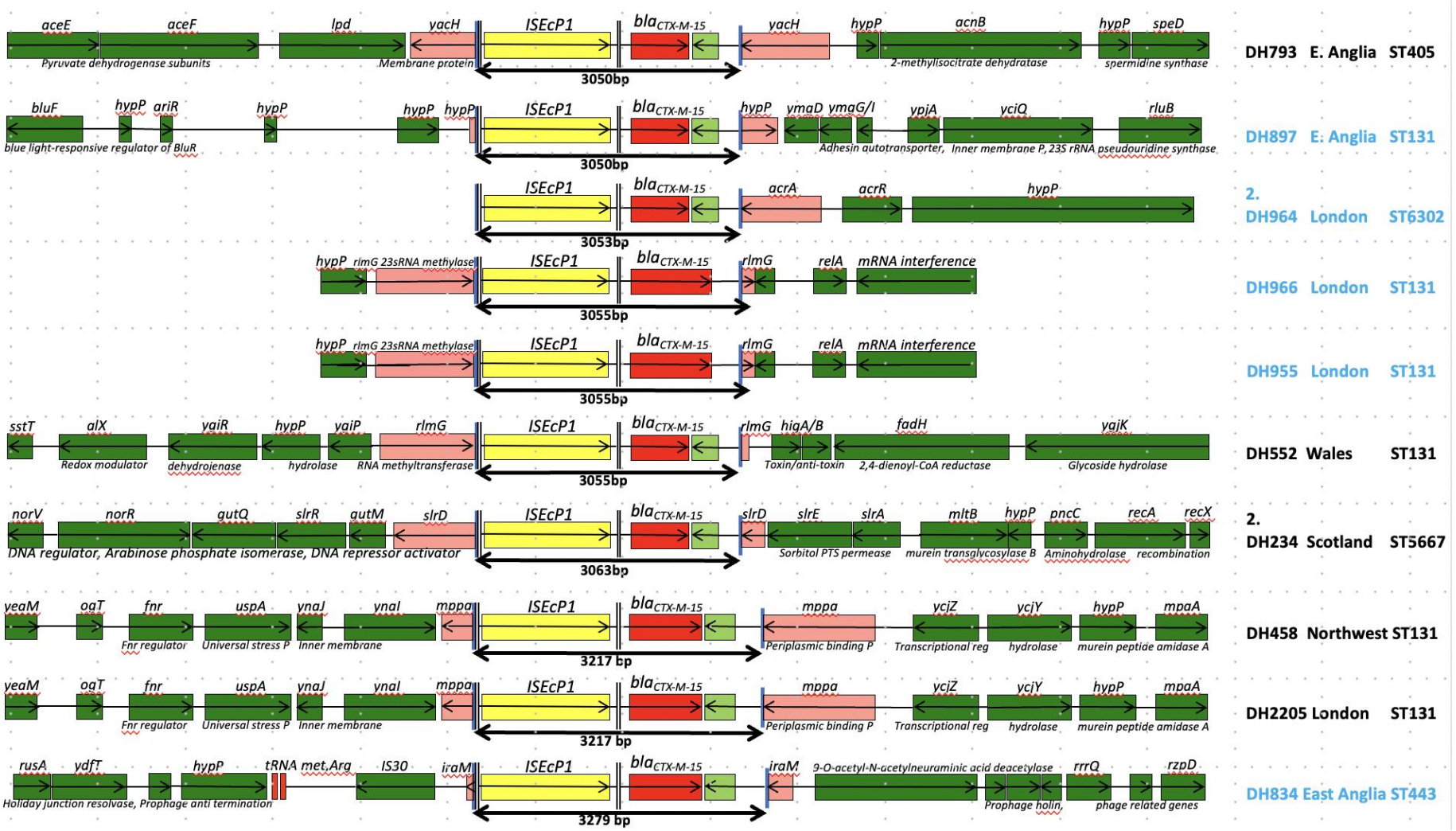


Figure 3.9.4. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

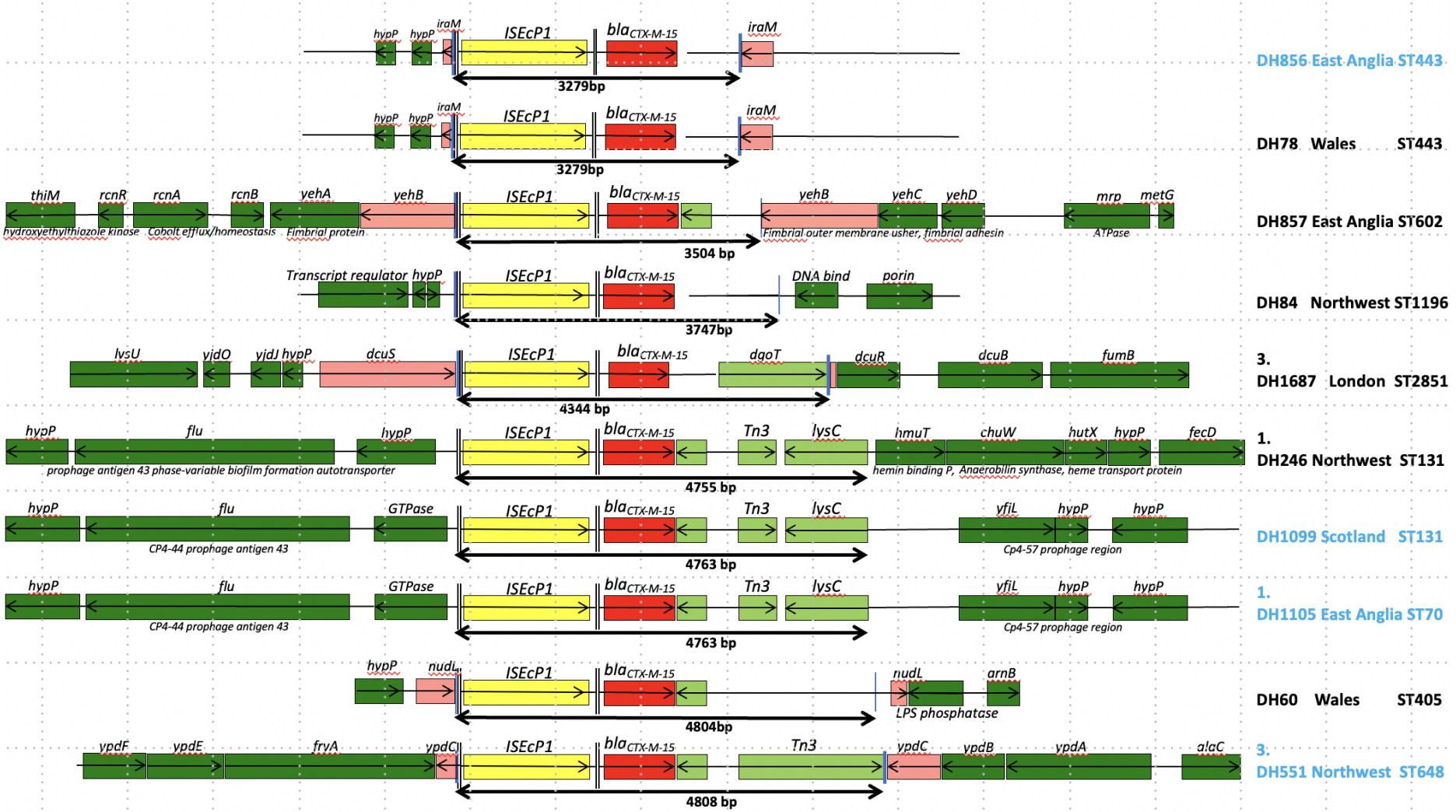


Figure 3.9.5. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

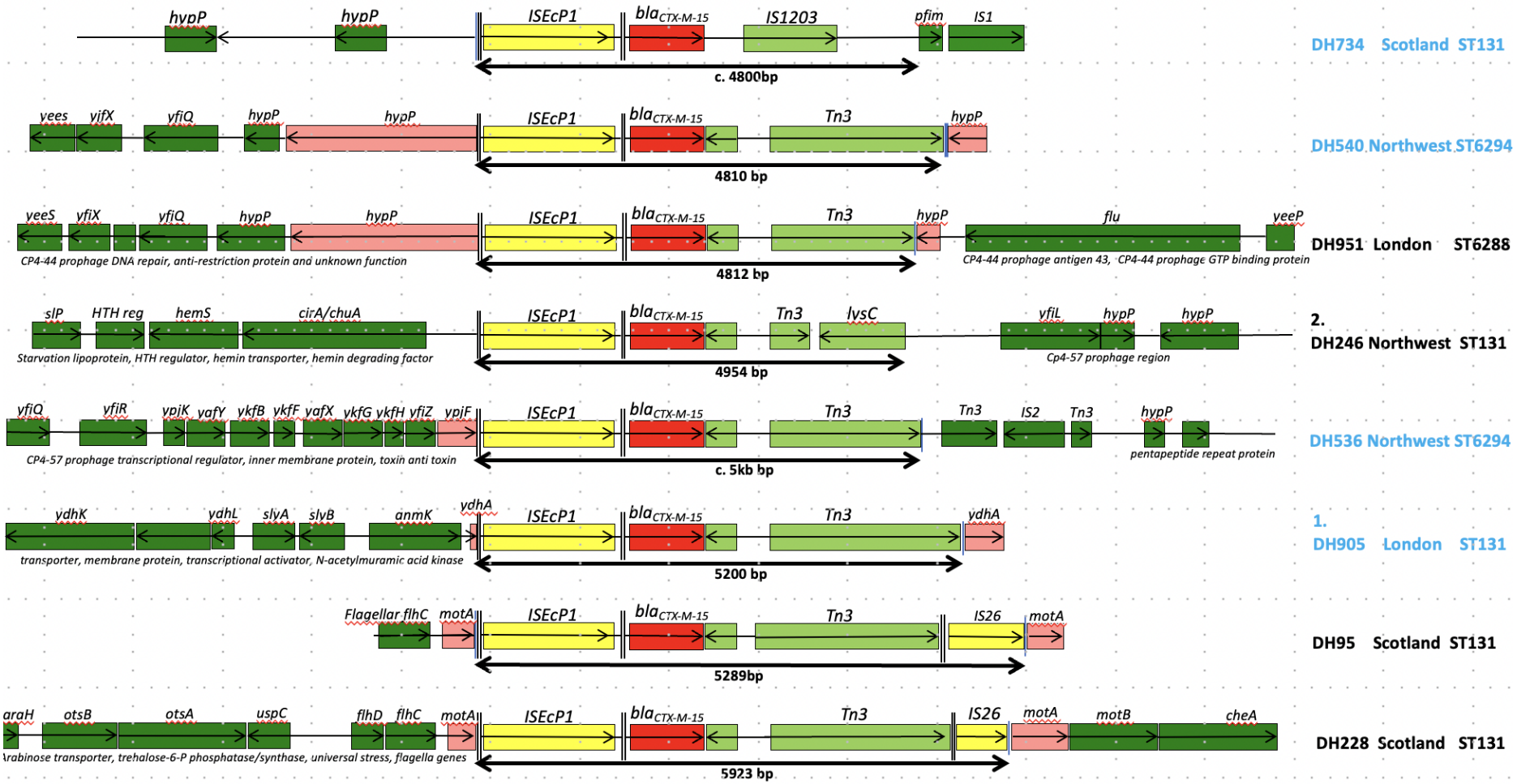


Figure 3.9.6. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*CTX-M-15 gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

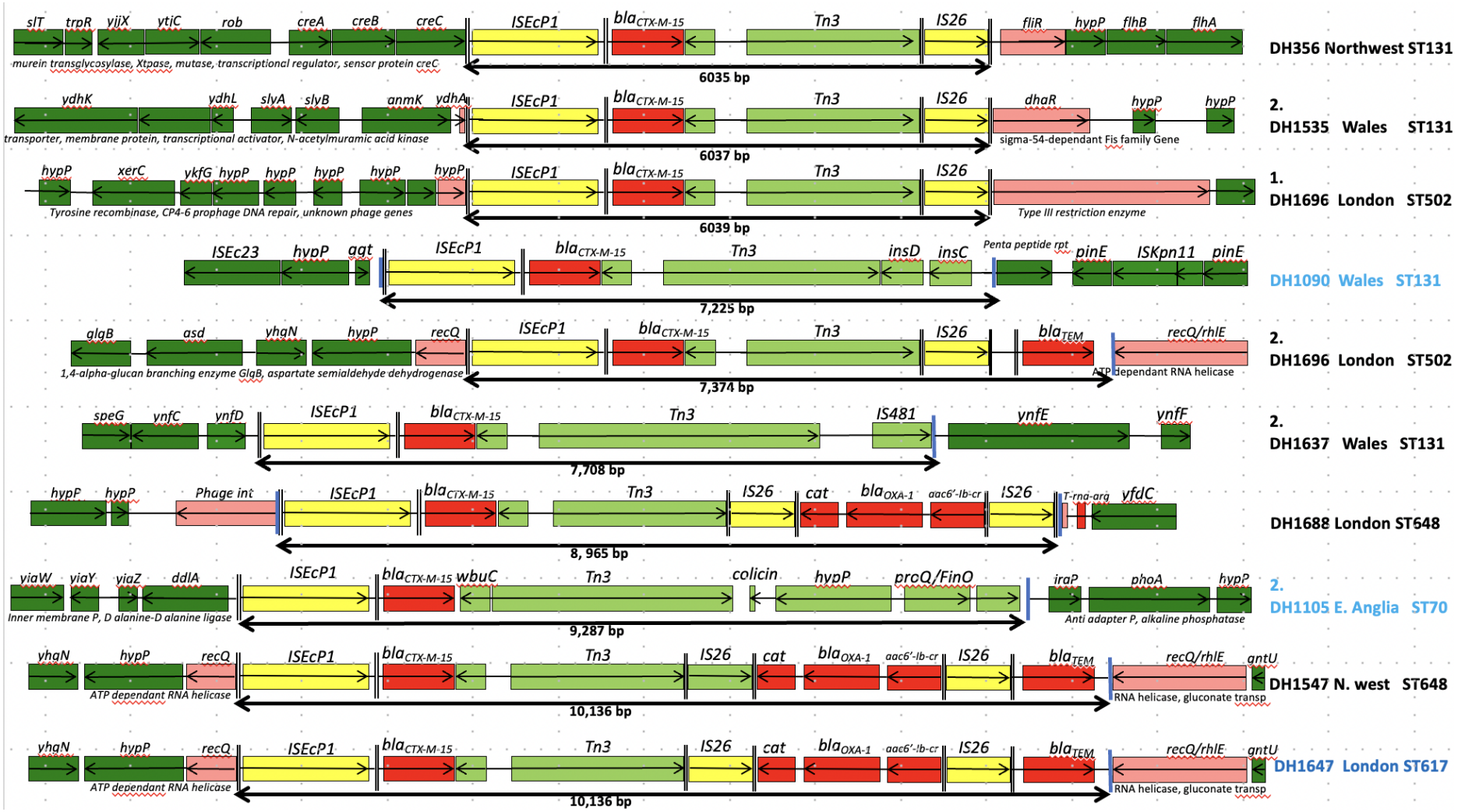


Figure 3.9.7. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

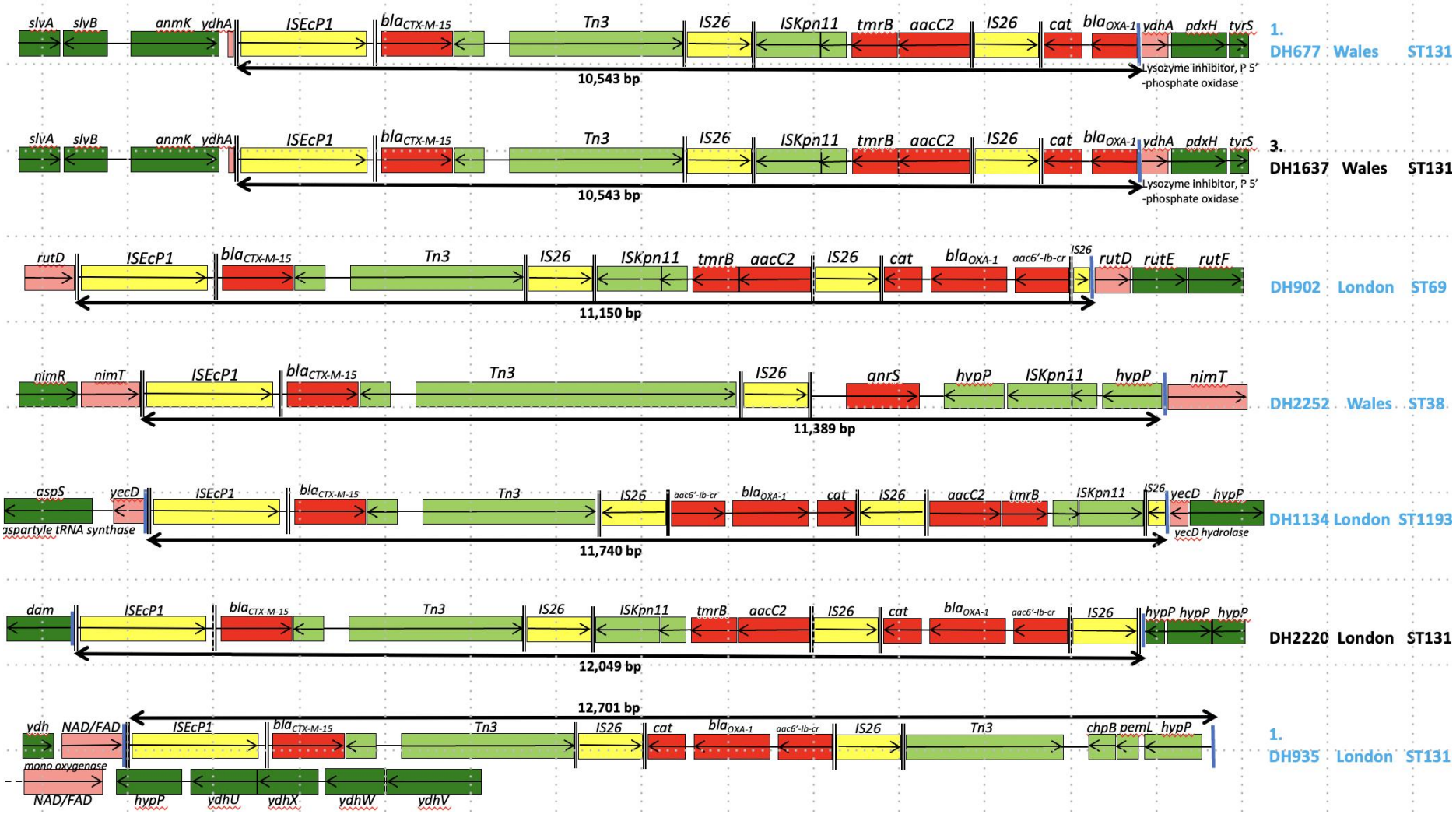


Figure 3.9.8. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

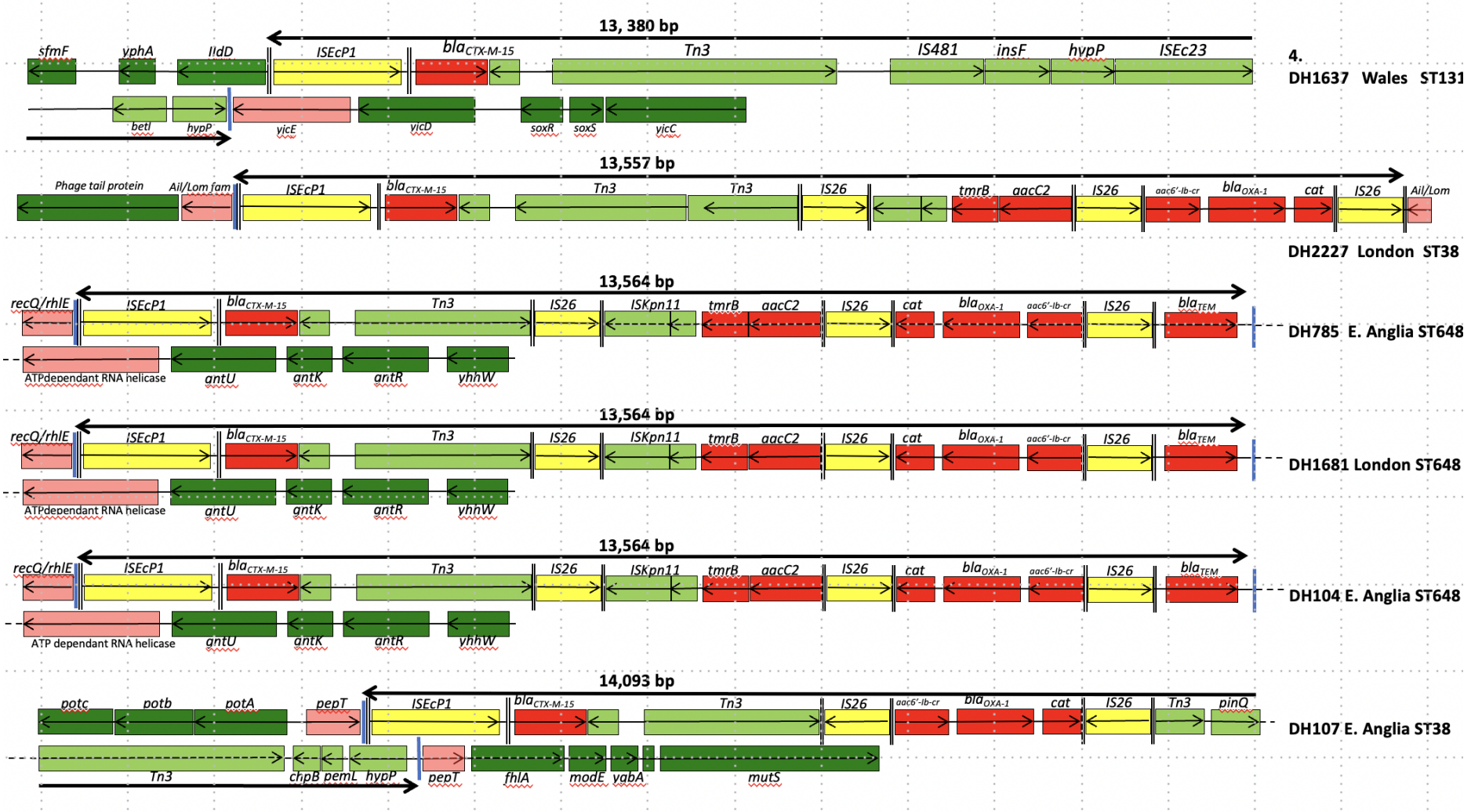


Figure 3.9.9. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*CTX-M-15 gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

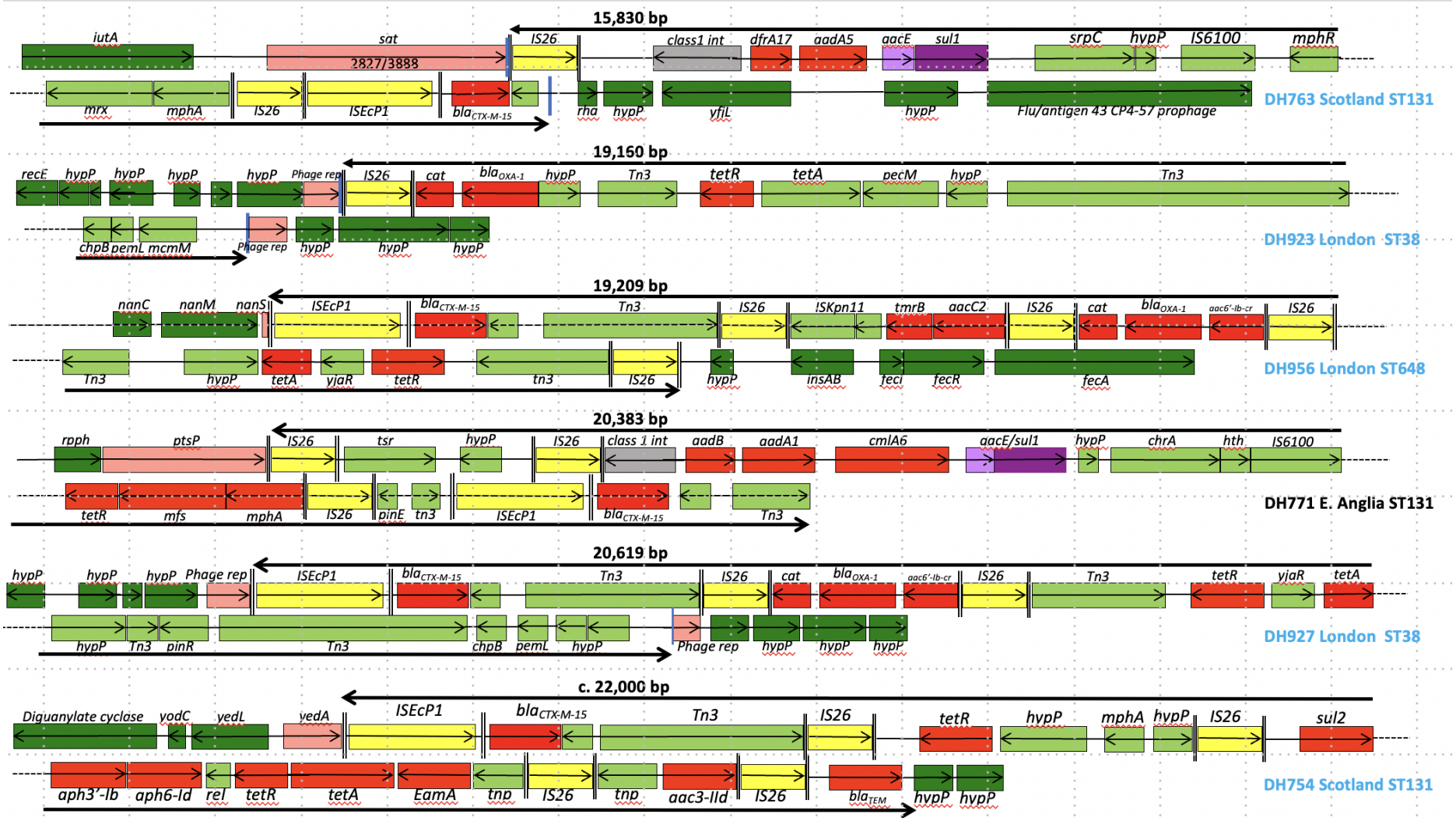


Figure 3.9.10. IS*EcP1* mediated chromosomal insertions.

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

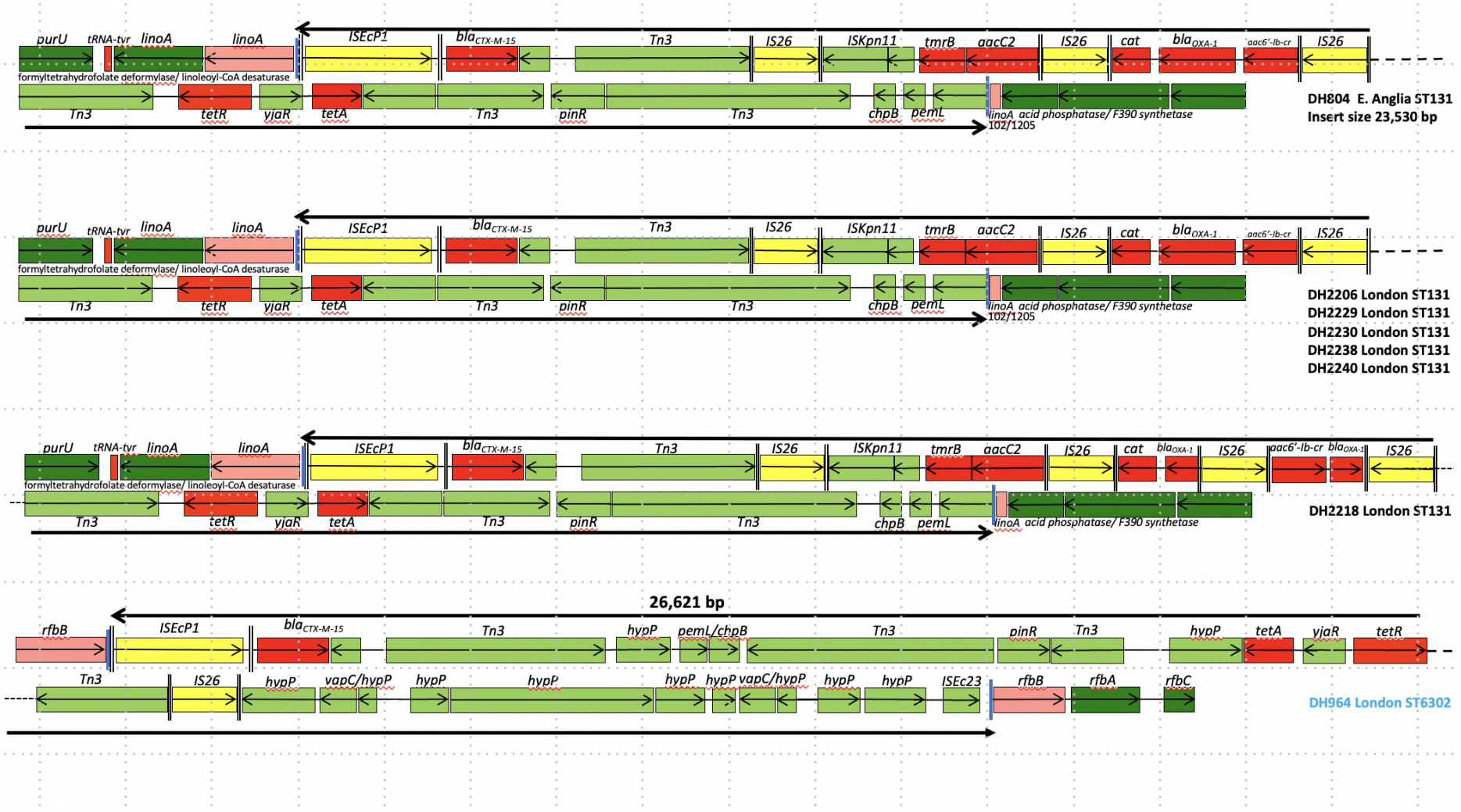


Figure 3.9.11. IS*EcP1* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

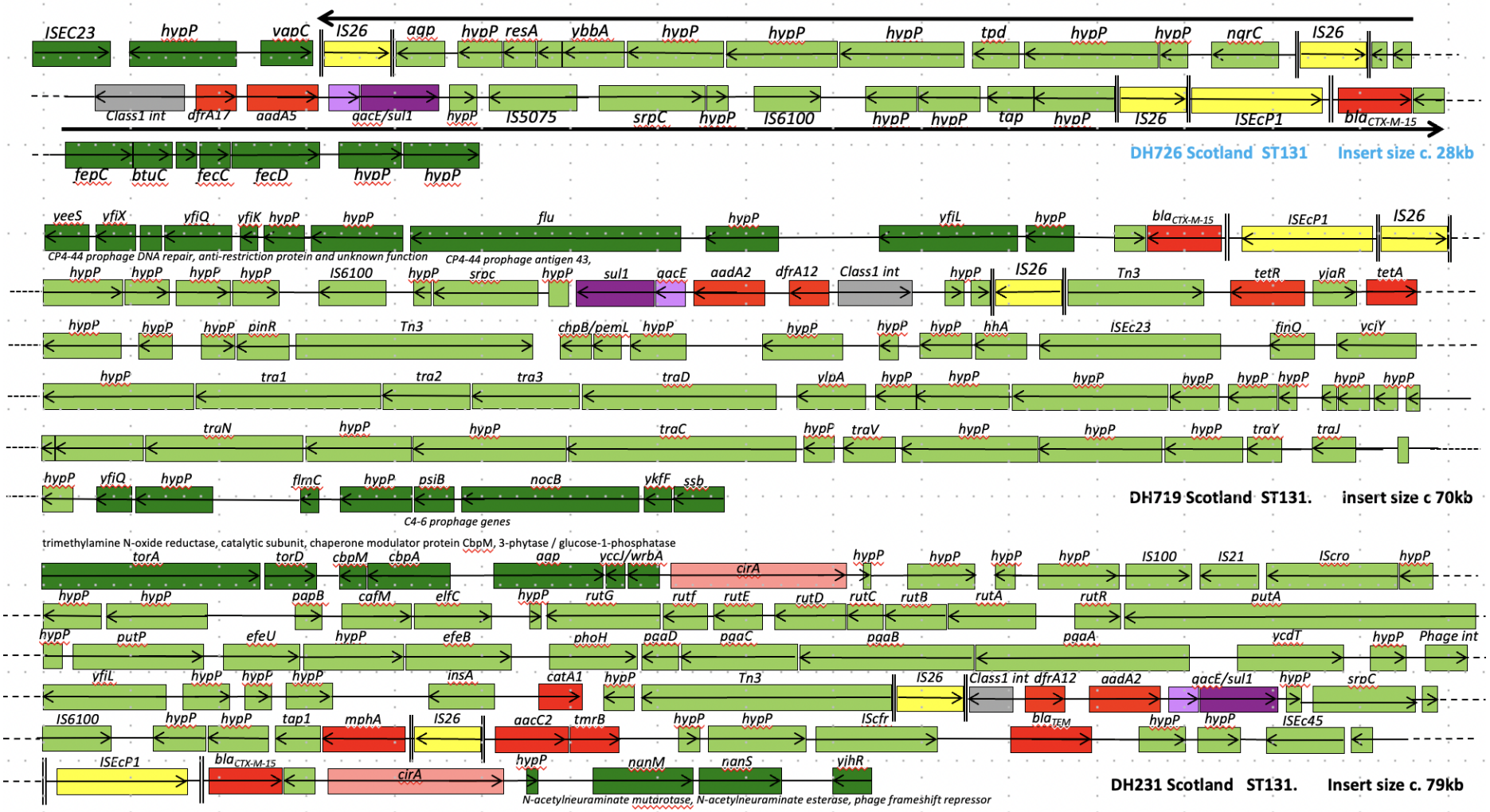


Figure 3.10. Large chromosomal insertions of complex construction

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

3.3.6 IS*26* mediated chromosomal insertions

Several chromosomal insertions were mediated by the IS*26* element, typically by two IS*26* elements functioning as a composite transposon mobilising large sections of DNA found between these elements. IS*26* was responsible for 30 chromosomal insertion events, 9 of these were associated with a target-site duplication event evidencing transposition. In general, IS*26* mobilisation events were of a larger size than IS*Ec*P1 mobilisation events being above 20 kb and ranging in size from c.7kb to >60kb. This element appeared to be capable of making extended resistance islands often associated with several resistance genes including the b-lactamase gene $bla_{\text{OXA-1,}}$ the bifunctional aminoglycoside and quinolone resistance gene ***aac6'-Ib-cr*** and a truncated version of the chloramphenicol resistance gene catB3 as well as several *IS26* elements. The *IS26* insertions were less random than the *ISEcP*1-mediated insertions and several identical insertion sites were observed. The transposition signal was also more often lost than for the IS*EcP1*-mediated transposition events and homologous recombination appeared to play a larger role in constructing extended resistance gene islands including class 1 integrons. Secondary homologous recombination events were also evidenced by different truncated genes often being found at the terminus of the chromosomal insertions.

Table 3.6. IS*26* mediated chromosomal insertions listed by size of insert

SEPSIS FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TE M	Adjacent/Inserted Genes	Insert size	DRIVER	DIRECT REPEAT	INSERT SITE
DH95	ST131	Scotland	**3.11.1**	+	+		dapA/IS26/tn3/tn3/CTX-M-15/ISEcp1/IS26/ugpE	7,034 bp	IS26	ND	LHS has dapA gene encoding a dihydrodipicolinate synthase family protein and RHS has upgE encoding a diglycerol phosphate transporter. NB not normally found adjacent therefore likely recombination since insertion. Section is also part of a larger resistance island including OXA-1 see description below.
DH95	ST131	Scotland	**3.11.1**	+	+		yhjB/IS26/cat/OXA-1.aac6'-Ib-cr/IS26/hypP/tetR/ yajR/hypP/tn3/pinR/tn3/chpB/pemL/hypP/tn3/IS26/ mcpQ/hypP/IS26/ehaG	20,673 bp	IS26	ND	Insertion is flanked by yhjB on One side encoding a putative DNA binding response regulator and ehaG on the other side encoding a autotransporter adhesin EhaG. Interestingly yhjC and yhjB are adjacent genes on the chromosome of AP026518 and therefore looks like the entire region containing these resistance islands and the intervening DNA has been acquired on mass. This whole section is of 147,559bp in length.
DH739	ST131	Scotland	**3.11.1**	+	+		ybfF/IS26/ISEcP-1/CTX-M-15/tn3/tn3/IS26/fabG	7,035 bp	IS26	ND	LHS has ybfF gene and RHS has fabG encoding a 3-oxo-acyl (acyl carrier protein) reductase. NB not normally found adjacent therefore likely recombination since insertion.
DH2223	ST131	London	**3.11.1**	+	+		molR/DR/IS26/aac6'-ib-cr/OXA-1/catB3/IS26/CTX-M15/hypP/tn3/IS26/tn3/IS26/DR/molR	10,361 bp	IS26	GAAATAC	IS26 mediated large insertion into molR gene at position 1032/3762 bp identical to several other insertions above.
DH422	ST131	Wales	**3.11.1**	+	+		molRmetreg/IS26/DISEcp1/DIS26/CTXM15/tn3/tn3/ IS26/ISKpn11/tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/ IS25/wgr domain	11,153 bp	IS26	ND	IS26 mediated insertion into molR metabolism regul;ator gene
DH2130	ST131	London	**3.11.1**	+	+		molRmetreg/IS26/DISEcp1/DIS26/CTXM15/tn3/tn3/ IS26/ISKpn11/tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/ IS25/wgr domain	11,153 bp	IS26	ND	
DH777	ST131	E. Anglia	**3.11.1**	+	+		molR/IS26/CTX-M-15/Tn3/IS26/tn3/IS26/ISKpn11/tmrB /aacC2/IS26/hypP/OXA-1/aac6'-Ib-cr/IS26/molbdate met reg/metG	12,325 bp	IS26	ND	IS26 mobilised insertion into moly region of the chromosome. NB there is a truncation event subsequent to insertion that is probably IS26 mediated (via homologous recombination) such that a section of 2013bp is missing relative to the complete sequence consisting of about half of molR and molybdate metreg genes. Again this is in the fimbrial operon.
DH806	ST131	E. Anglia	**3.11.1**	+	+		molR/IS26/CTX-M-15/hypP/Tn3/IS26/tn3/IS26/hypP/tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/molR	12,610 bp	IS26	GTATTTC	IS26 mediated insertion into molR at same position as several above.
DH803	ST131	E. Anglia	**3.11.1**	+	+		molR/IS26/CTX-M-15/hypP/Tn3/IS26/tn3/IS26/hypP/tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/molR	12,610 bp	IS26	GTATTTC	IS26 mediated insertion into molR at same position as several above.
DH2222	ST131	London	**3.11.1**	+	+		hypP/IS26/hypP/CTXM15/tn3/IS26/tn3/IS26/ISKPn11 /tmrB/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/metG	12,951 bp	IS26	ND	IS26 mediated large insertion with a hyP on the LHE inserted at position 396/2004 and metG at the RHE inserted at position 1669/2043 ie identical to several others above. No direct repeat sequence indicating that homologous recombination based on IS26 has occurred subsequent to insertion. Notably another IS26 is found adjacent to yehQ upstream in this strain. metG gene methionyl-tRNA synthetase.
DH2221	ST131	London	**3.11.1**	+	+		metG/IS26/aac6'-ib-cr/OXA-1/cat/IS26/aacC2/tmrB /ISKPn11/IS26/tn3/IS26/tn3/hypP/CTX-M15/hypP/IS26 /yehQ	12,951 bp	IS26	ND	IS26 mediated large insertion into metG encoding methionyl-tRNA synthetase at position 1659/2043 bp at the LHE and 416/ 2004 bp of a hypothetical protein close to yehQ at the RHE. These are well space on the AP026518 (894651/912393) genome indicating a secondary recombination event subsequent to insertion therefore no directly repeated sequence at either end.
DH905	ST131	London	**3.11.2**	+	+		**2.** molR/IS26/CTX-M-15/Tn3/IS26/Tn3/IS26/ISKpn11 /tmrB/aacC2/IS26/CAT/OXA-1/aac6'-Ib-cr/IS26/molR	12, 958 bp	IS26	TATTTC	**2.** Insertion is into molR family trans regulator gene at position 1022/3762 bp in a fimbriae region and close to fim genes metG and outer membrane usher genes and similar to several other insertions.

SEPSIS FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TE M	Adjacent/Inserted Genes	Insert size	DRIVER	DIRECT REPEAT	INSERT SITE
DH115	ST131	N. west	**3.11.2**	+	+		**molR/IS26/CTX-M-15/Tn3/IS26/tn3/IS26/ISKpn11/tmrB/aacC2/IS26/hypP/OXA-1/aac6'-Ib-cr/IS26/molR**	12,967 bp	IS26	GAATAC	IS26 mediated insertion of 12,967bp into molR gene atPosition 1032/3762bp
DH1647	ST131	London	**3.11.2**	+	+	+	**2. molR/IS26/CTX-M-15/Tn3/IS26/tn3/IS26/ISKpn11/tmrB/hypP/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/molR**	12,958 bp	IS26	GAATAC	**2.** IS26 mediated Insert into molR gene exact position as above ie 1025/3762bp.
DH716	ST131	Scotland	**3.11.2**	+	+		**molR/IS26/CTX-M-15/Tn3/IS26/tn3/IS26/ISKpn11/aacC2/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/molR**	12, 959 bp	IS26	GAATAC	Insertion is into the molR family transcriptional Regulator like many other insertions above.
DH105	ST131	E. Anglia	**3.11.2**	+	+	+	**molR/IS26/CTX-M-15/Tn3/IS26/tn3/IS26/ISKpn11/tmrB/aacC2/IS26/hypP/OXA-1/aac6'-Ib-cr/IS26/molbdate met reg/metG**	12,967 bp	IS26	ND	IS26 mediated insertion Into molR and molyddate met reg gene in exact place as Above.
DH108	ST131	E. Anglia	**3.11.2**	+	+	+	**chiA/IS26/hypP/aac6'-Ib-cr/IS26/pine/tn3/ISEcP-1/CTX-M-15/tn3/ISKpn11/tnrB/hypP/IS26/yhjC**	13,894 bp	IS26	ND	IS26 mediated insertion into position 1734/2703bp chiA gene encoding lysozyme. RHS is inserted upstream of yjhC encoding a putative HTH transcriptional regulator.
DH784	ST131	E. Anglia	**3.11.2**	+	+		**molR/IS26/CTX-M-15/Tn3/IS26/tn3/IS26/ISKpn11/tmrB/aacC2/IS26/catB3/OXA-1/aac6'-Ib-cr/IS26/molR?**	14,620 bp	IS26	ND	IS26 mediated inserted into molR family transporter gene insertion is 14965bp.
DH726	ST131	Scotland	**3.11.3**	+			**nqrC/IS26/agP1/hypP/resA/hypP/ybbA1/hypPx3/tpd/hypPx2/nqrC/IS2/hypP/xerD/xerC/ant1/qacE/sul1/hypP/IS110/srpC/hypP/IS6100/tap/hypP/IS26/ISEcP1/CTX-M-15/fepC**	16,901 bp	ISEcP1/IS26	ND	CTX-M-15 found in a resistance island Closely associated with ISEcp-1 so it probably got onto the chromosome via this element but subsequently the LH terminus (76bp) ofISEcp-1 has been truncated by IS26. The region includes Several plasmid genes and is similar to CP054215. Appears that IS26 mobilised section inserted into the ISEcp1/ctxM prior insertion
DH233	ST131	Scotland	**3.11.3**	+	+	-	**1. yhjC/IS26/aacC2/tmrB/ISKpn11/IS26/tn3/CTX-M-15/ISECP1/IS26/hypP/tap/hypPx2/IS6100/srpC/hypP/sul1/qacE/bcr1/ant/aadB/class1 int/IS26/yieK**	23,211 bp	IS26	ND	insertion into intergenic region adjacent to a HTH transcriptional regulator gene yhjC. NB another IS26 bracketted section is found in this genome encoding OXA-1 and other genes similar to that found in DH95 above. The whole section is of about 150kb (not drawn).
DH2202	ST131	London	**3.11.3**	+	+		**yqhC/IS26/aacC2/hypPx2/ISKPn11/IS26/tn3/CTX-M-15/hypPx4/pine/IS26/hypPx2/tap1/hypPx2/IS6100/srpC/hypP/sul1/qacE/cmlA1/aadA5/aadB/class1/IS26/upaG**	23,211 bp	IS26	ND	IS26 mediated insertion int chromosomal genes yqhC at position 681/900 bp at LHE of the inserted region and upaG position 3531/4344 bp at the RHE. yghC gene encodes a DNA binding transcriptional activator and upaG gene encodes an autotransporter adhesin UpaG. No direct repeat is observed and inserted genes are far apart on the chromosome of AP026518 (1940496-2566334) indicating a recombination event subsequent to insertion in this strain.
DH1705	ST131	London	**3.11.3**	+	+	+	**yqhC/IS26/tn3/CTX-M15/ISEcP1/tn3/hypP/IS26/class1/aadB/aadA5/cmlA1/sul1/qac/srpc/hypP/IS6100/hypP/mphA/IS26/tap/IS26/aac6'-Ib-cr/Oxa-1/cat/IS26/yhjB**	24,550 bp	IS26	ND	
DH1540	ST131	Wales	**3.11.4**	+	+	+	**yqhC/IS26/tn3/CTX-M15/ISEcP1/hypP/tap/hypP/hypP/IS6100/hypP/srpc/sul1/qacE/cmlA1/aadA5/aadB/class1/IS26/tn3/IS26/aac6'-Ib-cr/Oxa-1/cat/IS26/yhjB**	24,584 bp	IS26	ND	
DH776	ST131	E. Anglia	**3.11.4**	+	+		**nqrC/IS26/hypP/tap/hypPx2/IS6100/hypP/srpc/hypP/sul/qacE/aadA5/dfrA17/class1/CTXM15/tn3/tn3/IS26/cat/Oxa-1/aac6'-Ib-cr/IS26/tn3/tetR/yjaR/tetA/hypPx3/pinR/tn3/tn3/chpB/pemL/hypP**	c. 28kb	IS26	ND	This insertion has similarities to DH726 in that it appears that IS26 section mobilised into another chromosomal site with various antibiotic resistance genes. Actual IS26 section is 20,127 bp long and includes class1 integron.

SEPSIS FAECAL ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OX A-1	TE M	Adjacent/Inserted Genes	Insert size	DRIVER	DIRECT REPEAT	INSERT SITE
DH839	ST131	E. Anglia	**3.11.4**	+	+		**Beta-gluc/IS26/aac6'-Ib-cr/OXA-1/cat/IS26/aacC2/ tmrB/ISKpn11/IS26/tn3/IS26/tn3/hypP/CTX-M-15/IS26 /hypPx4/yehL/yehM/hypPx3/yehQ/hypPx3/ypdA2/ mirA/yehW/yehX/osmF/ /IS26/molR/ beta-gluc**	37,595 bp	IS26	ND	IS26 mediated insertion into position 996/2298 of beta glucosidase Gene. The RHE of the insertion is into the exact site of the molR gene as Described in several isolates above. There appears to have been A large invertion event since the other end of the beta glucosidase Gene is at the opposite end of IS26 as would been observed with A direct insertion. Perhaps insertion is based on homologous recombination of beta gal sequence.
DH804	ST131	N. West	**3.11.4**	+	+		**MolR/IS26/yehA1/yehA2/yehB2/yehB3/yehC/yehD/y ehE/mrp/metG/hypPx2/IS26/gatZ/gatY3/fbaB/yegT/ yegU1/yegU2/yegV/yegW/yegX/thiD/thiM/rcnA/IS26/ CTX-M-15/tn3/IS26/tn3/IS26/ISKpn11/tmrB/IS26/cat/OXA-1/aac6'-Ib-cr/IS26/gatZ**	24,365 bp 39,900 bp overall	IS26	ND	IS26 mobilised insertion into position 633/1263 of gatZ gene encoding D-tagatose-1-6-bisphosphate aldolase subunit GATZ which includes the CTX-M-15 gene. This appears to be part of a larger resistance island inserted into the exact same position of molR gene as those described above. There also appear to be several homologous recombination Events that have occurred such that this section now contains A section of the fim operon etc.
DH1709	ST131	Wales	**3.11.5**	+	+		**IS26/class1/dfrA17/aadA5/qac/sul/hypPx2/chrA/tra ns/hypPx3/tap2/hypP/IS26/sul2/hypP/IS26/hypPx3/I S1182/IS26/ISEcP1/CTX-M-15/hypP/tn3/IS26/cat/Oxa-1/aac6'-Ib-cr/IS26/tn3/tetR/yajR/tetA/hypPx2/pinR /tn3/chpB/pemL/hypP/ISPu7/ISEc12.**	>40kb	UNCLEAR Both-IS26 &ISEcP1	ND	Large region (c.40kb-perhaps bigger-maybe worth some more investigation?) found on several plasmids but missing plasmid transfer region. includes many IS26 elements and ISEcp1. Can't identify exact ends of the chromosomal insertion but many chromosomal genes on the same fragment. Upstream is a large pro-phage region similar to that found in AP026518.
DH95	ST131	Scotland	**3.11.5**	+	+		**dapA/IS26/Tn3/ CTX-M-15/ISEcp1/IS26/16kb blu light resp reg/IS26/tn3/hypP/pemL/chbB/tn3/pinR/tn3/ hypP tetA/yajR/tetR/tn3/IS26/aac6'-Ib-cr/ OXA-1/ cat/ IS26/ aacC2/tmrB/ISKpn11/IS26/dapA**	43,985 bp	IS26	GTTCA GCC	dapA gene encodingA dihydrodipicolinate synthase gene at position 615/888bp .
DH510	ST131	N. west	**3.11.5**	+	+		**1. gatZ/IS26/aac6'-Ib-cr/OXA-1/CAT/IS26/hypP/IS26/ yehA/yehB/yehC/yehD/yehE/mrp/metG/hypP/hypP/ ISKpN11/tmrB/aacC2/IS26/gatZ/gatY/fbaB/yegT/yeg U/yegV/yegW/yegY/thiD/thiM/hypP/rcnR/rcnA / rcnB/IS26/CTX-M-15/tn3/IS26/tn3/IS26/hypP/ hypP/IS26/mglC2** **2. IS26/CTX-M-15/tn3**	c. 44kb c.6kb	IS26	ND	Inserted region appears to be produced by two separate insertion events Firstly an insertion of aac6'/OXA-1/cat/ plus a fimbrial operon Section and tmbR/aacC2 into a gatZ gene encoding a tagatose-1-6-bisphosphate aldolase 2 protein. This insertion is of size 23,029bp and is bounded by IS26 elements with a direct repeat of AACGCCTC indicating a transposition event.Then there appears a Second recombination based insertion based on IS26 Sequence including CTX-M-15.This includes several IS26 Elements and is of size 9767bp. It appears to be IS26Mediated although no direct evidence of transposition So may have been homologous recombination based or subsequent . This insertion is 11,677bp from the OXA-1 insertion. There is a second CTX-M-15 insertion but is complex and about 6kb in length and not obvious how it got there. It is close to rcn gene which on AP genome Is close to a membrane usher.
DH976	ST131	London	**3.11.5**	+	+		**mglC/IS26/tn3/IS26/tn3/ctx-M-15/IS26/rcnB/rcnA/rcnR /thiM/thiD/ yegX/ yegW/yegV/yegT/fbaB/gatZ/IS26/ aacC2/tmrB/ISKpn11/IS26/hypP/metG/mrp/yehE/ye hD/yehC/yehB/yehA1/IS26/hypPx3/IS26/cat/aac6'-Ib-cr/ OXA-1 /IS26/gatZ**	44,502 bp	IS26	ND	IS26 mediated. The left hand end is located At position 887/1011bp of mglC encoding a galactoside ABC transporter permease MglC.
DH1680	ST131	London	**3.11.5**				**molR/DR/IS26/tfaE/sftE/pine/aroA/serC/hypP/ycaO/ hypP/pfiB/pfiA/ycaD/ycaC/dmsC/dmsB/dmsA/xerC/ hypPx2/parM/hypPx20/pinE/IS26/IS26/aac6'-Ib-cr/OXA-1/catB3/IS26/aacC2/tmrB/ISKpn11/IS26/tn3/hypP/C TXM15/IS26/DR/molR**	64,281 bp	IS26	AATAC	IS26 mediated insertion into molR at same site as others above. Insertion includes many plasmid genes and resembles a hfr strain.

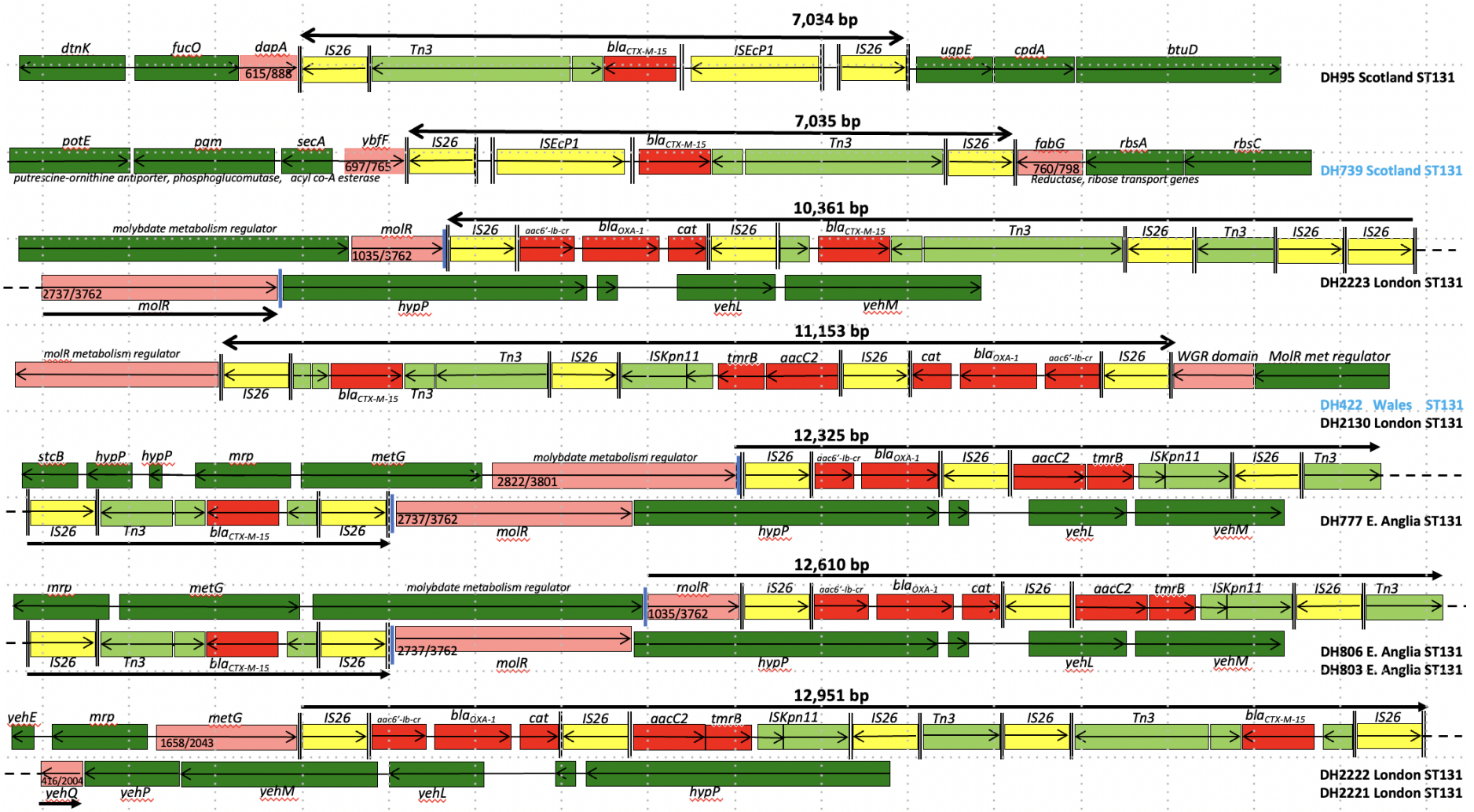


Figure 3.11.1. *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the bla$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeat flanking the insertion site. Isolate name in black and blue indicate bacteraemia and faecal isolates, respectively.

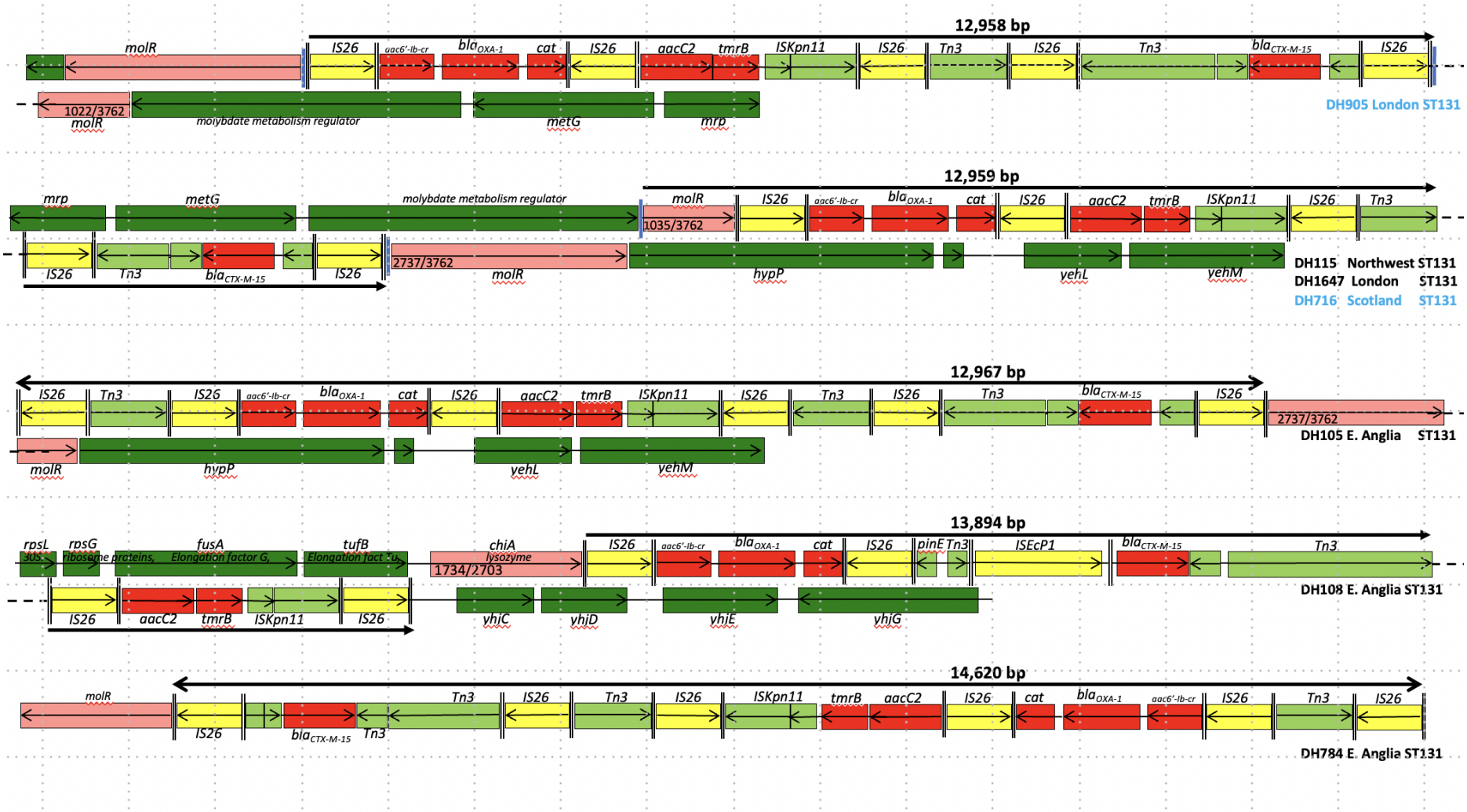


Figure 3.11.2. *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the $bla_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

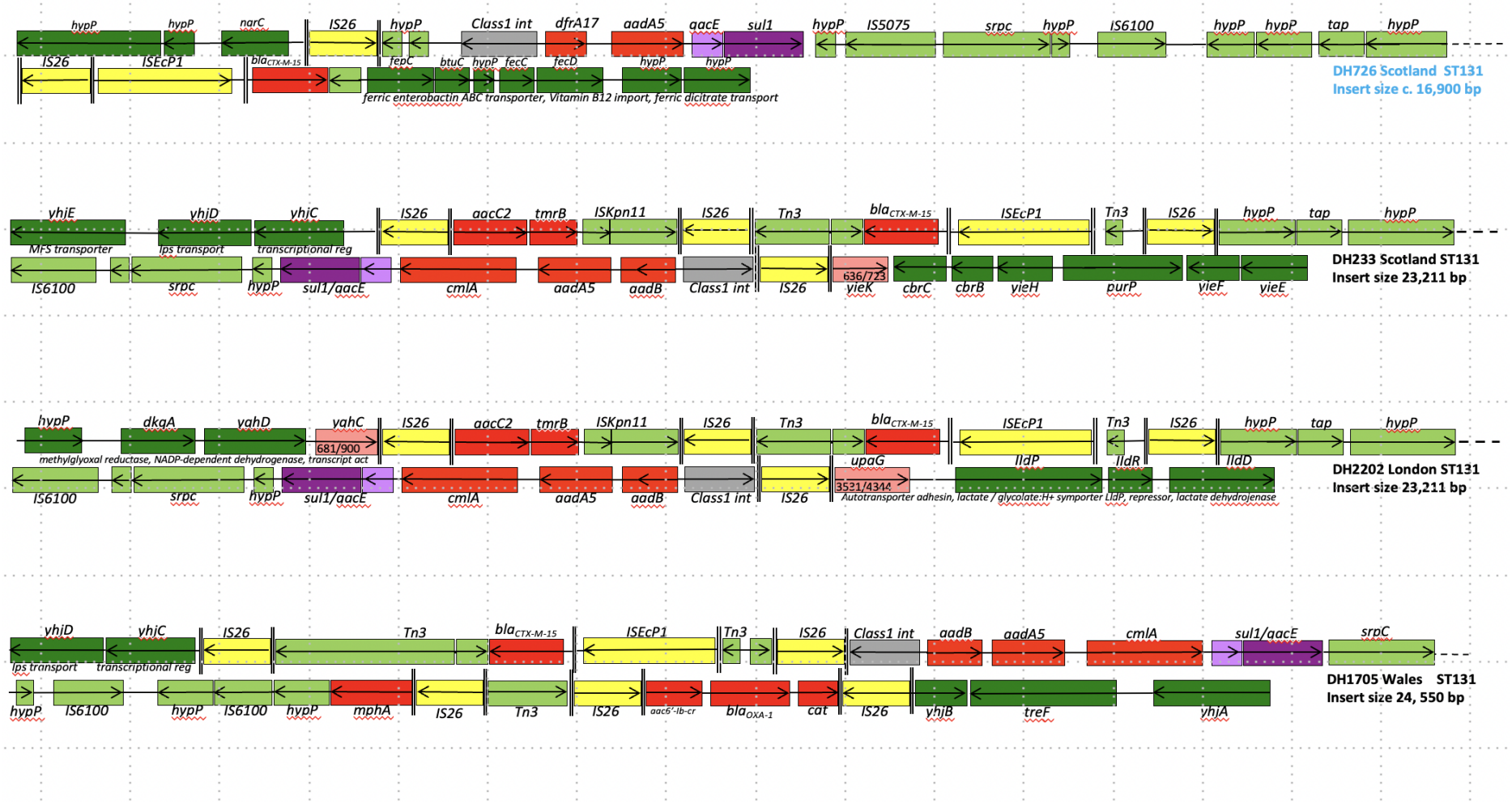


Figure 3.11.3. *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeat flanking the insertion sites. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

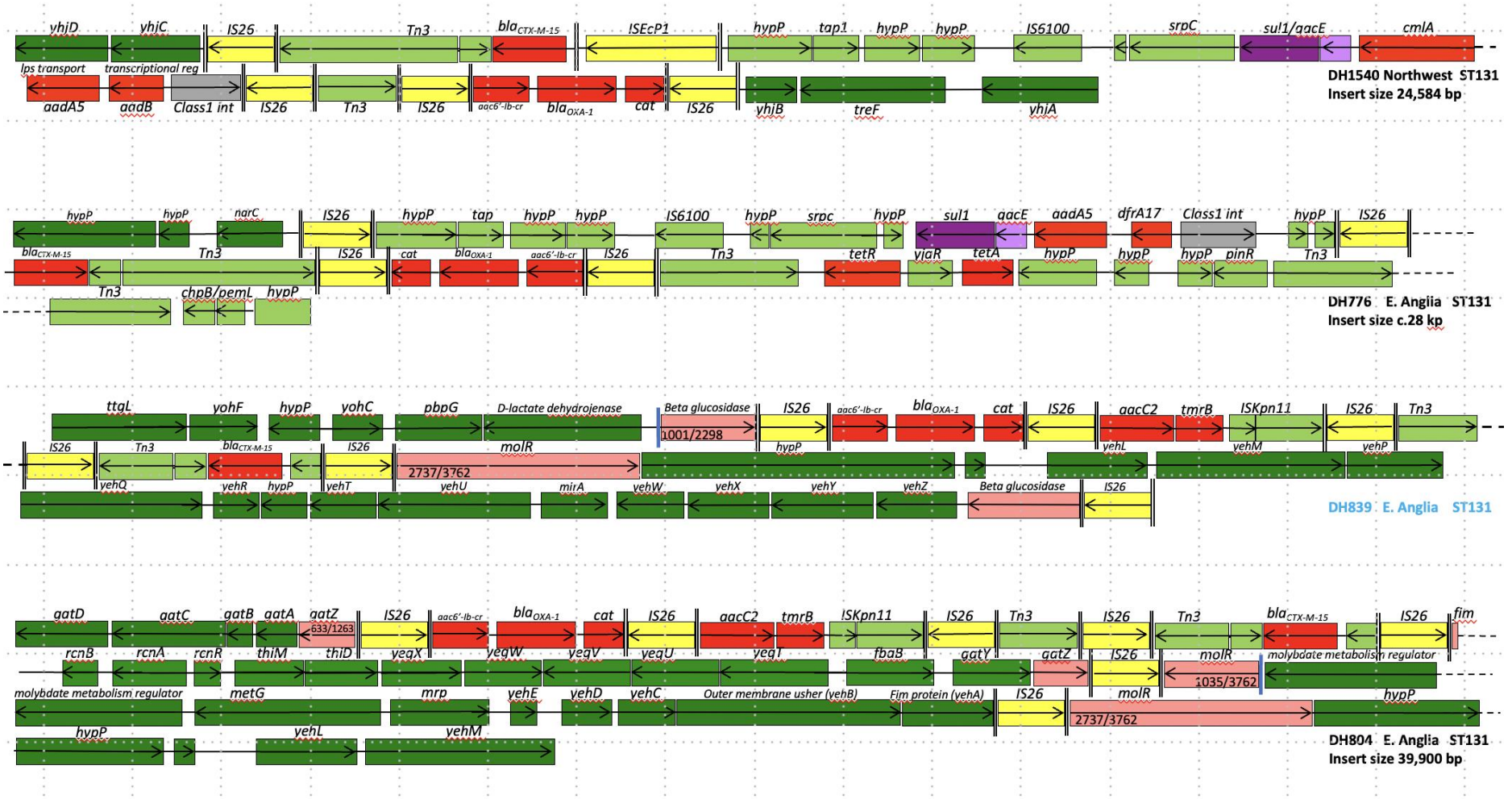


Figure 3.11.4. *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

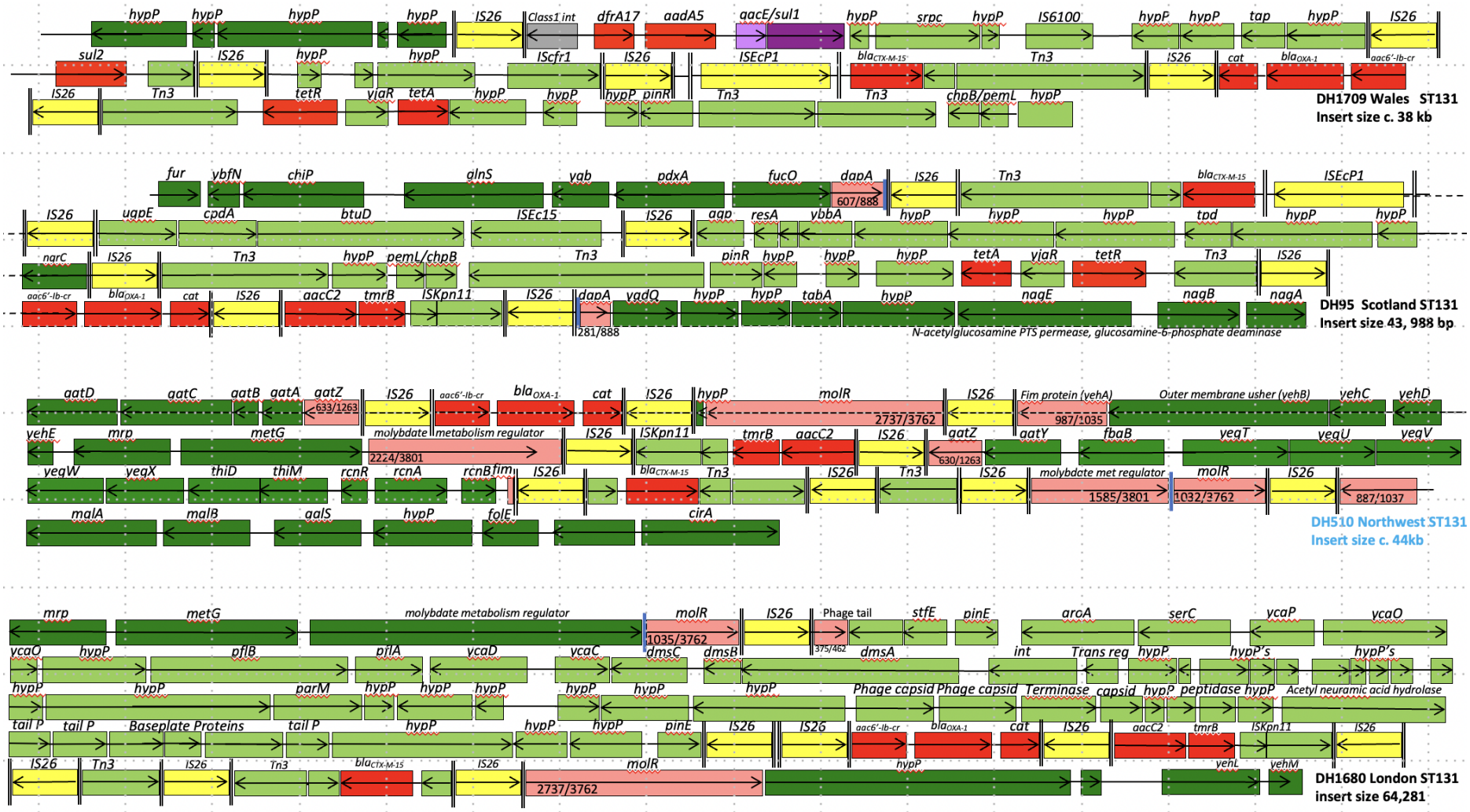


Figure 3.12. Complex *IS26* chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeat flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

3.3.7 Effects on the *E. coli* chromosome

Chromosomal insertions mediated by both *ISEc*P1 and *IS*26 often caused damage to chromosomal genes. In silico analysis of chromosomal integration sites indicated that intergenic insertions were surprisingly rare. We found less than a dozen chromosomal integrations that did not cause direct gene damage by inserting into intergenic spaces between structural genes (**Table 3.7**). In contrast, we found nearly 100 integrations that destroyed resident chromosomal genes forming pseudogenes upon insertion (**Table 3.8**). There were some obvious hotspots for insertion in our collection of isolates. This was notably within an identical position in a *molR* gene encoding a molR family transcriptional regulator which was IS26 mediated and found in 18 ST131 isolates from all regions of the UK and both from bacteraemia and faecal sources (**Table 3.8**). Similarly, we found a hot spot of insertion for *ISEc*P1 in the *recQ* gene encoding an ATP-dependant DNA helicase of ST648 isolates derived from bacteraemia's and for *ISEc*P1 in the *linoA* gene encoding a linoleoyl-coA desaturase of ST131 isolates derived from bacteraemia's in London (**Table 3.8**). Other hot spots were *pepT* encoding a peptidase T and *mppa* encoding a peptide ABC transporter substrate-binding protein.

Chromosomal insertions were also a substrate for chromosomal rearrangements such as invertion and translocation events mediated by homologous recombination based on repetitive sequence as evidenced by the numerous insertions that were found to have different pseudogenes found at the termini of the inserts (**Table 3.9**). Despite having several hot-spots for insertion around the chromosome there seems to be a general random nature to the chromosomal insertions as indicated in (**Figure 3.13**) with insertions randomly

arranged around the entire chromosome with perhaps increased insertion within prophage regions (**Figure 3.13**).

Table 3.7 Chromosomal intergenic insertion sites

LHS gene ID	function	LHS gene Location in reference genome	RHS gene ID	Adjacent in ref genome	function	RHS gene Location in reference genome	Isolate ID	origin
yhjC 603 bp	helix-turn-helix transcriptional regulator	2,463,160 -> 2,464,041	*yhjB* 723 bp	YES	helix-turn-helix transcriptional regulator 603 bp BDP14697.1	2,462,622 -> 2,462,020	DH1705 (ST131) DH1540 (ST131)	Bacteraemia (W) Bacteraemia (NW)
uspC 429 bp	universal stress protein BDP12948.1	571,551 -> 571,979	*flhD* 351 bp	YES	flagellar transcriptional regulator FlhD	569,572 -> 569,222	DH212 (ST131)	Bacteraemia (S)
gstB 627 bp	glutathione S-transferase GstB BDP16465.1	4,423,184 -> 4,422,558	*dacC (Pbp)* 1212 bp	YES	D-alanyl-D-alanine carboxypeptidase BDP16466.1	4,423,422 -> 4,424,633	DH551 (ST648)	Faecal (NW)
hypP 657 bp	hypothetical protein BDP14079.1	1,827,799 -> 1,827,143	*cirA* Length: 2091 bp	YES	ligand-gated channel BDP14078.1	1,826,643 -> 1,824,553	DH551 (ST648) DH941(ST3075)	Faecal (NW) Faecal (L)
ynfD 306 bp	membrane protein BDP12645.1	263,309 -> 263,614	*ynfE* 2427 bp	YES	dimethyl sulfoxide reductase subunit A BDP12646.1	263,813 -> 266,239	DH1637 (ST131)	Bacteraemia (W)
ddlA 1095 bp	D-alanine--D-alanine ligase BDP16043.1	3,963,767 -> 3,962,673	*iraP* 261 bp	YES	anti-adapter protein BDP16044.1	3,964,230 -> 3,964,490	DH1105 (ST70)	Faecal (EA)
Dam 765 bp	restriction endonuclease subunit BDP16616.1	4,565,143 -> 4,565,907	hypP 216 bp	YES	hypothetical protein BDP16615.1	4,564,737 -> 4,564,952	DH2220 (ST131)	Bacteraemia (L)
INTER GENE CHROMOSOMAL INSERTIONS IN PROPHAGE REGIONS								
yfjL 1491 bp	hypothetical protein BDP14100.1	1,847,154 -> 1,845,664	GTPase 873 bp	YES	hypothetical protein BDP14101.	1,848,486 -> 1,849,358	DH1105 (ST70) DH1099 ST131)	Faecal (EA) Faecal (S)
chuA CDS 1713 bp	product: ligand-gated channel protein protein_id: BDP14678.1	2,439,768 -> 2,438,056	*yfjL* 816 bp	YES	product: hemin ABC transporter substrate-binding protein protein_id: BDP14680.1	2,440,821 -> 2,441,636	DH246 (ST131)	Bacteraemia (NW)
sat 3888	autotransporter BDP14065.1	1,809,624 -> 1,813,511	Rha 207 bp	YES	Rha family transcriptional regulator CDS BDP14098.1	1,844,679 -> 1,844,885	DH763 (ST131)	Faecal (S)
papI 222 bp	major pilus subunit operon regulatory protein protein_id: BDP14052.1	1,793,296 -> 1,793,517	transposase 801 bp	YES intergenic	transposase protein_id: BDP14053.1	1,794,968 -> 1,794,168	DH956 (ST648)	Faecal(L)

Table 3.8. Pseudogenes formed by chromosomal insertion events

Gene name	function	Gene length	PROTEIN ID	Location in reference genome	Isolate ID	Isolate origin	Isolate location
MolR	MolR family transcriptional regulator protein_id: BDP13266.1	3762 bp	BDP13266.1	898,990 -> 902,751	DH716 (ST131) DH806 (ST131) DH739 (ST131) DH105 (ST131) DH112 (ST131) DH803 (ST131) DH115 (ST131) DH839 (ST131) DH1647 (ST131) DH1680 (131) DH784 (131) DH905 (131) DH422 (131) DH777 (131) DH804 (131) DH510 (131) DH2130 (131) DH2223 (131)	Faecal Bacteraemia Faecal Bacteraemia Bacteraemia Bacteraemia Bacteraemia Faecal Faecal Bacteraemia Bacteraemia Faecal Faecal Bacteraemia Bacteraemia Faecal Bacteraemia Bacteraemia	Scotland E. Anglia Scotland E. Anglia Northwest E. Anglia Northwest E. Anglia London London E. Anglia London Wales East Anglia East Anglia Northwest London London
recQ	ATP-dependent DNA helicase	2097	BDP14605.1	2,369,001 -> 2,366,905	DH1547 (ST648) DH1688 (ST648) DH785 (ST648) DH1681 (ST648) DH104 (ST648) DH1696 (ST648)	Bacteraemia Bacteraemia Bacteraemia Bacteraemia Bacteraemia Bacteraemia	Northwest London E. Anglia London E. Anglia Northwest
linoA	linoleoyl-CoA desaturase	1078	BDP17020.1	4,945,580 -> 4,946,657	DH804 (ST131) DH2206 (ST131) DH2218 (ST131) DH2229 (ST131) DH2230 (ST131) DH2238 (ST131) DH2240 (ST131)	Bacteraemia Bacteraemia Bacteraemia Bacteraemia Bacteraemia Bacteraemia Bacteraemia	E. Anglia London London London London London London
pepT	peptidase T	1227	BDP16832.1	4,779,579 -> 4,780,805	DH107 (ST38) DH1688 (ST648) DH1707 (ST394)	Bacteraemia Bacteraemia Bacteraemia	E. Anglia London Wales
mppA	peptide ABC transporter substrate-binding protein	1614	BDP12512.1	107,285 -> 108,898	DH458 (ST131) DH782 (ST131)	Bacteraemia Bacteraemia	Northwest E. Anglia

					DH2205 (ST131)	Bacteraemia	London
yhjC	helix-turn-helix transcriptional regulator	603	BDP14698.1	2,463,160 -> 2,464,041	DH233 (ST131) DH1540 (ST131) DH1705 (ST131)	Bacteraemia Bacteraemia Bacteraemia	Scotland Northwest Wales
iraM	anti-adapter protein	336	BDP17070.1	4,990,003 -> 4,989,668	DH78 (ST443) DH834 (ST443) DH856 (ST443)	Bacteraemia Faecal Faecal	Wales East Anglia East Anglia
nrfG	heme lyase NrfEFG subunit NrfG	597	BDP15272.1	3,124,467 -> 3,125,063	DH964 (ST6302)	Faecal	London
ydhA	lysozyme inhibitor	324	BDP12702.1	323,306 -> 322,983	DH677 (ST131) DH1637 (ST131) DH905 (ST131) DH1535 (ST131)	Bacteraemia Bacteraemia Faecal Bacteraemia	Wales Wales London Wales
FAD/NAD(P)	FAD/NAD(P) binding domain-containing protein	1605	BDP12728.1	347,658 -> 349,262	DH935 (ST1310	Faecal	London
Hydrolase	hydrolase	567	BDP12923.1	544,456 -> 545,022	DH1134 (1193)	Faecal	London
MotA	flagellar motor protein	888	BDP12944.1	568,514 -> 567,627	DH95 (ST131) DH228 (ST131)	Bacteraemia Bacteraemia	Scotland Scotland
rfbB	NAD(P)-dependent oxidoreductase	900	BDP13192.1	811,061 -> 810,162	DH964 (ST6302)	Faecal	London
gatZ	D-tagatose-1,6-bisphosphate aldolase subunit	1263	BDP13245.1	875,104 -> 873,842	DH112 (ST131) DH839 (ST131)	Bacteraemia Faecal	Scotland E. Anglia
beta-glucosidase CDS	beta-glucosidase	2298	BDP13282.1	924,848 -> 922,551	DH839 (ST131	Faecal	E. Anglia
PtsP	phosphoenolpyruvate--protein phosphotransferase PtsP	2247	BDP13921.1	1,652,713 -> 1,650,467	DH771 (ST131)	Bacteraemia	E. Anglia
dcuA	anaerobic C4-dicarboxylate transporter DcuA	1302	BDP15346.1	3,214,140 -> 3,212,839	DH1535 (ST131)	Bacteraemia	Wales
rfaH	transcription antitermination protein RfaH	489	BDP15030.1	2,838,328 -> 2,837,840	DH551 (ST648)	Faecal	Northwest
bluF	diguanylate phosphodiesterase	1212	BDP16905.1	4,843,612 -> 4,842,401	DH520 (ST131)	Faecal	Northwest

ynfK	ATP-dependent dethiobiotin synthetase BioD	696	BDP12653.1	272,111 -> 271,416	DH1687 (ST2851)	Bacteraemia	London
artM	arginine ABC transporter permease protein ArtM	669	BDP16487.1	4,443,993 -> 4,443,325	DH1637 (ST131)	Bacteraemia	Wales
pckA	phosphoenolpyruvate carboxykinase [ATP]	1623	BDP14572.1	2,323,954 -> 2,325,576	DH1707 (ST394)	Bacteraemia	Wales
cfa	cyclopropane-fatty-acyl-phospholipid synthase	1149	BDP12723.1	341,939 -> 343,087	DH1707 (ST394)	Bacteraemia	Wales
htrE	outer membrane usher protein	2601	BDP15809.1	3,704,644 -> 3,702,044	DH810 (ST636) DH1098 (ST636) DH960 (ST636)	Bacteraemia Faecal Faecal	Northwest London London
rbsR	ribose operon repressor	993	BDP14938.1	2,730,761 -> 2,731,753	DH234 (ST5667)	Bacteraemia	Scotland
yjfO	hypothetical protein	330	BDP15394.1	3,263,242 -> 3,262,913	DH722 (ST652)	Faecal	Scotland
srlC/D	PTS sorbitol transporter subunit IIA	372	BDP13763.1	1,469,438 -> 1,469,809	DH826 (ST5667) DH234 (ST5667)	Faecal bacteraemia	East Anglia Scotland
dgoT	MFS transporter	1293	BDP14877.1	2,664,459 -> 2,663,167	DH1687 (ST2851)	Bacteraemia	London
yehB	fimbrial outer membrane usher protein	2481	BDP13259.1	889,612 -> 887,132	DH857 (ST602)	Bacteraemia	East Anglia
yacH	hypothetical protein	1473	BDP15786.1	3,680,029 -> 3,678,557	DH793 (ST405)	Bacteraemia	East Anglia
nudL	putative Nudix hydrolase NudL	579	BDP12866.1	488,415 -> 488,993	DH60 (ST405)	Bacteraemia	Wales
dcuS	sensor histidine kinase DcuS	1527	BDP15325.1	3,185,553 -> 3,184,027	DH1687 (ST2851)	Bacteraemia	London
ypdC	AraC family transcriptional regulator CDS	858	BDP13519.1	1,193,469 -> 1,194,326	DH551 (ST648)	Faecal	Northwest
yecD	hydrolase CDS	567	BDP12923.1	544,456 -> 545,022	DH1134 (ST1193)	Faecal	London
rlmA	ribonuclease R	2442	BDP15384.1	3,253,216 -> 3,255,657	DH2202 (ST394)	Bacteraemia	London
UvrY	DNA-binding response regulator	657	BDP13024.1	629,757 -> 629,101	DH456 (ST131)	Bacteraemia	Northwest

hypP	hypothetical protein	1740	BDP15537.1	3,407,202 -> 3,405,463	DH1705 (ST648)	Faecal	Northwest
acrF	acrEF/envCD operon transcriptional regulator	663	BDP14444.1	2,208,515 -> 2,207,853	DH825 (ST2851)	Faecal	East Anglia
acrA	MexE family multidrug efflux RND transporter periplasmic adaptor subunit	1086	BDP16123.1	4,048,138 -> 4,047,053	DH964 (ST6302)	Faecal	London
rutD	putative aminoacrylate hydrolase RutD	801	BDP16721.1	4,669,020 -> 4,668,220	DH902 (ST69)	Faecal	London
dapA	dihydrodipicolinate synthase family protein	888	BDP16314.1	4,253,821 -> 4,252,934	DH95	Bacteraemia	Scotland
g*atZ*	D-tagatose-1,6-bisphosphate aldolase subunit	1263	BDP13245.1	875,104 -> 873,842	DH510 (ST510	Faecal	Northwest
nanS	putative 9-O-acetyl-N-acetylneuraminic acid deacetylase	1008	BDP15567.1	3,437,480 -> 3,436,473	DH956 (ST648)	Faecal	London
PSEUDOGENES GENERATED IN PROPHAGE REGIONS BY CHROMOSOMAL INSERTION EVENTS							
transposase	transposase	801	BDP14053.1	1,794,968 -> 1,794,168	DH734 (ST131)	Faecal	Scotland
Phage Tail	tail protein CDS	396	BDP16883.1	4,821,890 -> 4,822,285 Bases:	DH212 (ST131)	Bacteraemia	Scotland
rlmG/ygjO	ribosomal RNA large subunit methyltransferase G	1137	BDP14276.1	2,038,701 -> 2,037,565	DH552 (ST3) DH955 (ST131) DH966 (ST131	Faecal Faecal Faecal	Northwest London London
hypP	outer membrane fluffing protein CDS flu gene	3123	BDP14102.1	1,849,692 -> 1,852,814	DH540 (ST6289) DH951 (ST6288)	Faecal Faecal	Northwest London
Tnsp (Pseudogene)	transposase	327bp		1,795,294 -> 1,794,968	DH734 (ST131)	Faecal	Scotland
ypjF	CP4-57 prophage; toxin of the YpjF-YfjZ toxin-antitoxin system	343		1,845,676 -> 1,845,461	DH536 (ST6294)	Faecal	Northwest
hypP	hypothetical protein	372	BDP16914.1	4,848,207 -> 4,848,578	DH897 (ST131)	Faecal	East Anglia

Ail/Lom	ND (Not found in reference genome)	DH2227
nimT	ND	DH2252
Phage integrase	ND	DH1688
Phage replication	ND	DH923, DH927
Cas	ND	DH97
hypP	ND	DH667

Table 3.9. Chromosomal insertions generating secondary chromosomal rearrangment events

LHS gene ID	function	LHS gene Location in reference genome	RHS gene ID	Adjacent in ref genome	function	RHS gene Location in reference genome	Isolate ID	origin
MolR 3762 bp *pseudogene*	MolR family transcriptional regulator BDP13266.1	898,990 -> 902,751	*metG* 2043 bp pseudogene	NO	methionine--tRNA ligase protein_id: BDP13264.1	892,994 -> 895,036	DH804 DH2222 DH2221	Bacteraemia (EA) Bacteraemia (L) Bacteraemia (L)
MolR 3762 bp *pseudogene*	MolR family transcriptional regulator BDP13266.1	898,990 -> 902,751	molybdate metabolism regulator 3801 bp *pseudogene*	YES	molybdate metabolism regulator	895,177 -> 898,977	DH422 (ST131) DH777 (ST131) DH2130 (ST131)	Faecal (W) Bacteraemia (EA) Bacteraemia (L)
yhjC 603 bp	helix-turn-helix transcriptional regulator	2,463,160 -> 2,464,041	*yieK* 723 bp *pseudogene*	NO	glucosamine-6-phosphate deaminase	2,692,024 -> 2,691,302	DH506 (ST131) DH233 (ST131)	Faecal (NW) Bacteraemia (S)
yhjC 603 bp	helix-turn-helix transcriptional regulator	2,463,160 -> 2,464,041	*ChiA* 723 bp *pseudogene*		Lysozyme *NB no equivalent in ref sequence*		DH108 (ST131)	Bacteraemia (EA)
Hly *NB nothing similar in reference sequence*			*dcuA* 1302 bp *pseudogene*	NO	anaerobic C4-dicarboxylate transporter BDP15346.1 **(Pseudogene)**	3,214,140 -> 3,212,839	DH1535 (ST131)	Bacteraemia (W)
nqrC 762 bp	Na(+)-translocating NADH-quinone reductase subunit C		*fepC* 810 bp	NO	ABC transporter	1,972,030 -> 1,971,221	DH726 (ST131)	Faecal (S) Bacteraemia (EA)

NB nothing similar in reference sequence							DH776 (ST131) LHS only	
yedA pseudogene	putative inner membrane transporter YedA BDP13069.1	665,249 -> 666,169	*Indistinct end*	NO			DH754 (ST131)	Faecal(S)
gatZ 1263 *pseudogene*	D-tagatose-1,6-bisphosphate aldolase subunit GatZ BDP13245.1	875,104 -> 873,842	*MglA 1011 bp pseudogene*	NO	*galactoside ABC transporter permease MglC BDP13297.1*		DH510 (ST131)	Faecal (NW)
gatZ 1263 *pseudogene*	D-tagatose-1,6-bisphosphate aldolase subunit GatZ BDP13245.1	875,104 -> 873,842	MolR 3762 bp *pseudogene*	NO	MolR family transcriptional regulator protein_id: BDP13266.1	898,990 -> 902,751	DH839 (ST131) DH804 (ST131	Faecal (EA) Bacteraemia (EA)
mglC 1011 bp *pseudogene*	galactoside ABC transporter permease MglC BDP13297.1	940,630 -> 939,620	fimbrial protein Length: 1035 bp *pseudogene*	NO	fimbrial protein protein_id: BDP13258.1	887,116 -> 886,082	DH976 (ST131	Faecal (L)
dapA 879 bp *pseudogene*	4-hydroxy-tetrahydrodipicolinate synthase BDP13599.1	1,280,424 -> 1,279,546	*ugpE* 846 bp	NO	sn-glycerol-3-phosphate transport system permease BDP14617.1	2,379,915 -> 2,379,070	DH95 (ST131)	Bacteraemia (S)
PROPHAGE GENES								
ydhA 324 bp *pseudogene*	lysozyme inhibitor	323,306 -> 322,983	*dhaR* 1842 bp *Pseudogene*	NO	sigma-54-dependent Fis family transcriptional regulator BDP16952.1	4,882,550 -> 4,884,391	DH1535 (ST131) DH1696 (Not exact but similar loc) (ST648)	Bacteraemia (W) Bacteraemia (NW)
ggt 1035 bp *pseudogene*	phosphotriesterase-related protein BDP13122.1	742,771 -> 743,805	RHS not found in reference genome	NO			DH1090 (ST131)	Faecal (W)
hypP 803 bp	hypothetical protein BDP14101.1	1,848,486 -> 1,849,358	*hmuT* 816 bp BDP14680.1	NO	hemin ABC transporter substrate-binding protein	2,440,821 -> 2,441,636	DH246 (ST131)	Bacteraemia (NW)

y*qhC* 900 bp *pseudogene*	AraC family transcriptional regulator BDP14190.1	1,941,119 -> 1,940,220	U*paG* 4344 bp *pseudogene*	NO	adhesin protein_id: BDP14781.1	2,562,803 -> 2,567,146	DH2202	Bacteraemia(L)
yfjL 1491 bp	hypothetical protein BDP14100.1	1,847,154 -> 1,845,664	*chuA* CDS 1713 bp *pseudogene*	NO	ligand-gated channel protein BDP14678.1	2,439,768 -> 2,438,056	DH246 (ST131)	Bacteraemia (NW)
fliR/lfiQ 273 bp *Pseudogene*	flagellar biosynthesis protein BDP15895.1	3,803,084 -> 3,802,812	*creC* 1425 bp	NO	two-component sensor histidine kinase BDP15656.1	3,531,009 -> 3,532,433	DH356 (ST131)	Bacteraemia (NW)
lldD 1017 Not found in reference sequence			*yjcE* 1650 bp *pseudogene*	NO	Na+/H+ antiporter BDP15259.1	3,110,364 -> 3,112,013	DH1637 (ST131)	Bacteraemia (W)
hypP DH1709 insertion c. 38kb not found in reference genome								

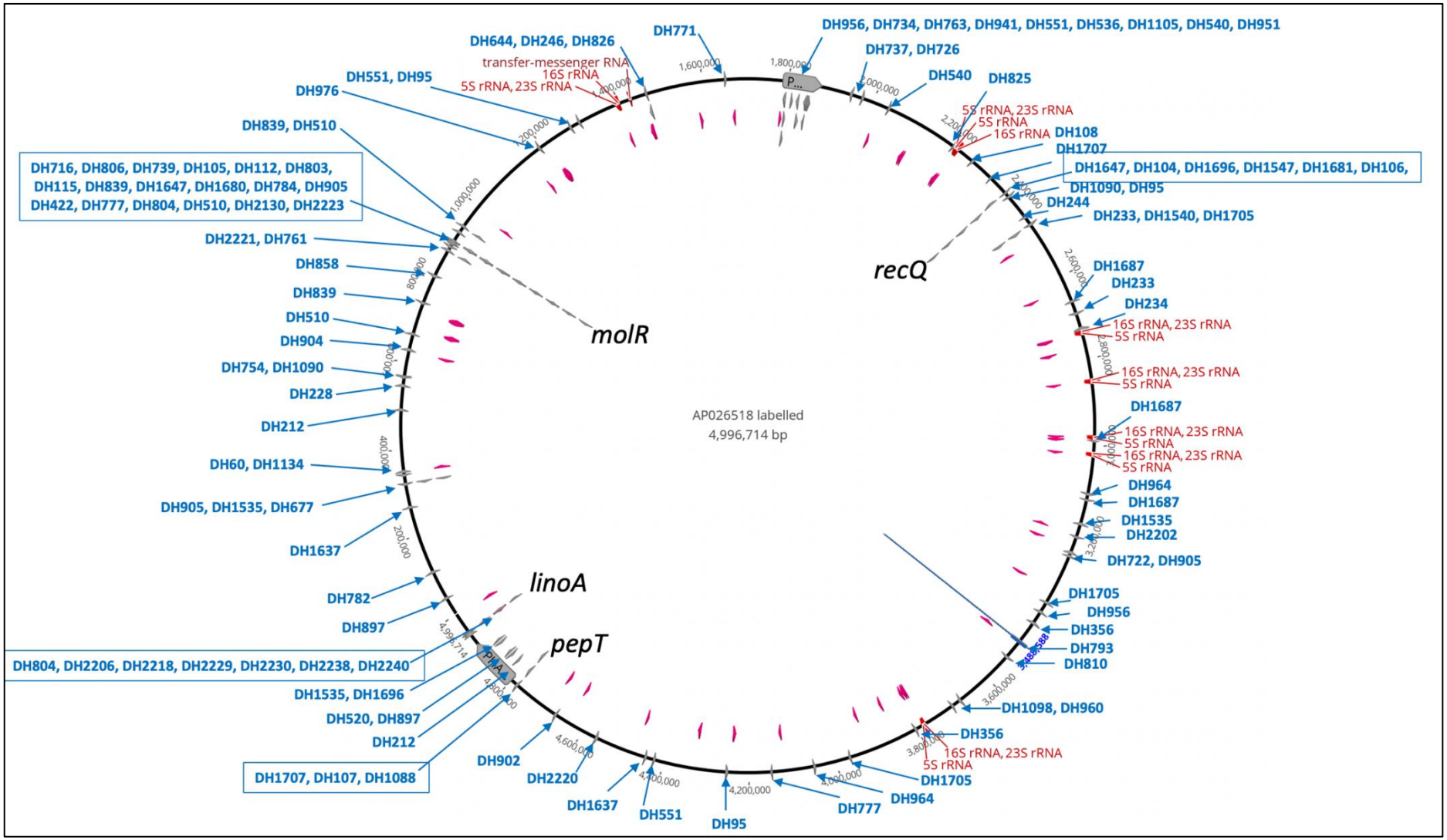


Figure 3.13. Location of insertions of *bla*$_{\text{CTX-M-15}}$ around the *E. coli* chromosome

All insertions are located with respect to the reference *E. coli* ST131 genome AP026518. Hot spots are indicated with bold gene identifiers and isolates having insertions at these hot are indicated with blue boxes. The location of two prophages are indicated with large grey arrows. The ribosomal operons are labelled in red.

3.4 Discussion

Chromosomal targeting of antibiotic resistance genes has been comprehensively studied in the laboratory over the last decade with a view to understanding the plasmid paradox. This is the idea that plasmids are essentially parasites that impose a fitness burden on their hosts such that they likely grow more slowly than plasmid-free variants. The advantage the host bacteria has is only seen in an environment containing antibiotics such that when the selection is removed the fitter non-resistant bacteria easily outcompete their parasitised neighbours, which is the fundamental basis of all prescription policies. To successfully compete, it has been hypothesised that antibiotic-resistance genes would naturally migrate to the chromosome and then the donating plasmid would be lost by the bacterial cell and plasmids would essentially die out. This has in fact been demonstrated by several laboratory-based experiments. However, the opposite truth of the paradox is that the plasmid has developed many mechanisms to ensure that it stays closely attached to its host. How the plasmid paradox plays out in the real world has not really been studied in any depth and most reports in the literature of chromosomal carriage of resistance genes are incidental (Brockhurst and Harrison 2022). The movement of resistance genes to the bacterial chromosome is of great concern to the clinical environment as many clinically relevant bacteria are susceptible to antibiotics. The movement to the chromosome could have the effect of permanently altering the genetic make-up of clinical bacteria thus making them increasingly difficult to treat. This has in fact happened with group B Streps where a bottleneck of selection due to tetracycline use in the early antibiotic era removed many strains leaving high-level chromosomal *tet* resistance in this species. The emergence of *tet* resistance in virulent strains is

now thought to be behind the emergence of disease towards the end of the last century (Da Cunha et al. 2014).

Here we have used a comprehensive collection of antibiotic-resistant *E. coli* collected from both serious disease as well as normal carriage representing all common *E. coli* types found in five areas of the UK. We initially addressed the chromosomal location using PFGE and radioactive probing of the *bla*$_{\text{CTX-M-15}}$ gene. This proved very useful because it made it reasonably easy to screen the entire collection for chromosomal and plasmid versions of the gene. The results were striking in that this study demonstrates that chromosomal carriage of *bla*$_{\text{CTX-M-15}}$ is hugely important, at least in the UK. We found chromosomal carriage was slightly more important in bacteraemia isolates than in faecal ones which may be due to selection imposed in the clinical environment or it may also reflect increased virulence in chromosomally carried resistance perhaps based on increased growth rates which might be expected in isolates that have lost plasmid versions of the gene. A chromosomal location was also noted in a large range of *E. coli* ST also indicating that movement to the chromosome is a common occurrence in *E. coli*. Another striking observation was that movement of *bla*$_{\text{CTX-M-15}}$ to the *E. coli* chromosome was closely related to loss of the donating plasmid. Only a minimal number of isolates possessed both plasmid and chromosomal versions of *bla*$_{\text{CTX-M-15.}}$ Another surprising observation was that several cell lines had lost the *bla*$_{\text{CTX-M-15}}$ gene entirely. All isolates had been sequenced by illumina Miseq short-read sequencing in Colindale before being sent to Cardiff as part of the original study and therefore the gene must have been lost either by short-term storage on dorset egg slopes or during growth for our experiments. This may be related to the high transposition rate of the *ISEc*P1 insertion sequence. We initially used the short-read

sequence data to analyse 50 genomes. This was partly successful for some of the simple *ISEc*P1-based insertions but often the genomes collapsed due to repetitive DNA. Since long-read sequencing using Minion has rapidly developed during this project allowing 96 barcoding and high-accuracy translation of FAST5 files we repeated the difficult genomes and several more besides so that we successfully located insertions in 100 isolates and managed to close several genomes. The results again are striking in that we did not expect so many resistance genes to be co-transferred with the *bla*$_{\text{CTX-M-15}}$ gene forming extensive resistance gene islands on the chromosome in some of the isolates. This is particularly concerning as many of these strains leave only the carbapenems as good treatment options. The combination of *bla*$_{\text{CTX-M-15}}$ and *bla*$_{\text{OXA-1}}$ is a particularly important combination since *bla*$_{\text{OXA-1}}$ can give inhibitor resistance and the movement of class 1 integrons to the *E. coli* chromosome is also of great concern. This is particularly important as these lines are probably permanently changed with all genes routinely handed down by vertical descent.

The *ISEc*P1 element was first discovered in 2000 and is hugely successful now being part of the normal flora of many people in the UK (Karim et al. 2001). Its success is mostly down to the fact that only one copy of the element is necessary for transposition of any gene that it manages to move beside. This is in stark contrast to IS*26* which needs two copies of the element to bracket a useful gene before it can be mobilised. The increased transposition is seen in our collection as multiple copies of IS*EcP1* were observed in 10% of the genomes. *ISEc*P1 was also more often responsible for mobilisation events than IS*26* and this is even though *ISEc*P1 emerged in Enterobacterales in recent history as opposed to IS*26* which was responsible for mobilising *shv* from the *Klebsiella pneumonia* chromosome at the

beginning of the antibiotic era. Our study also demonstrates that *ISEc*P1 can move large sections of DNA even if it has a bias to moving smaller sections. The *ISEc*P1 transposition events may also leave a more stable structure after transposition since only one copy of IS*EcP1* would make homologous recombination events less likely. Several hot spots for insertion on the *E. coli* genome were discovered in our study including *molR, linoA, recQ, pepT* and *mmpa*. It is unclear whether these represent actual hot-spots for insertion or amplification of successful variants. The close link between individual ST for the *molR* and *linoA* insertions (all ST131 and all ST648, respectively) may suggest what is happening here is amplification of successful clones which are perhaps successful due to chromosomal resistance. However, *pepT* is more interesting in that whereas the *molR* and *linoA* insertions are all at the exact same position within the gene and all isolates share the same ST. The *pepT* insertions represent different insertion sites within the same gene and also in different *E. coli* ST. This perhaps suggests that there is an advantage in having this gene knocked out. One of the interesting facts that has emerged from the sequencing era is that nearly all pathogens have smaller genomes than their closest commensal comparitors. Indicating that loss of genes is part of the pathogenesis process. The *pepT* gene may be a good candidate to perform gene knockouts and test virulence in an insect or animal model.

In summary, we have demonstrated that chromosomal carriage of antibiotic resistance in *E. coli* is hugely important in this large contemporary collection of UK *E. coli.* Future work might look at the importance of hot spots of insertion and the stability of the chromosomal carriage.

.

Chapter 4 Difference between *E. coli* Types and Associated Resistance Mechanisms in Disparate Nations

4.1 Introduction

4.1.1 Distribution of *E. coli* Phylogroups in different countries.

Previously our group has performed community carriage studies in Pakistan and Northeast India (Zahra et al. 2018; Paul et al. 2020) indicating that the specific types of *E. coli* and associated resistance mechanisms can be very different in different countries. In this study we sought to extend these studies to other nations with different climates and also to understand how the use of antibiotics can have an effect on *E. coli* types. We investigated *E. coli* that were selected with and without antibiotics at the same time and from the same samples. We chose two countries with drastically different climates, mostly because of the ease of getting samples. That is because during this project I had the opportunity to return to Saudi Arabia to collect samples from this country, specifically Unaizah in the Qassim region which has a hot climate. In a similar way, because my supervisor often visited Kazakhstan which has a particularly cold climate, we chose Karaganda in Kazakhstan. The COVID situation limited our collections and we had planned to do several collections in each country during the course of study. However, the COVID-19 crisis caused us to limit this study to single collections from each country that were collected prior to 2020.

4.1.2 Saudi Arabia

There is limited information on the types of *E. coli* that are carried in the community in Saudi Arabia and most of the studies that have been performed are based on urinary tract infections and antibiotic resistant isolates. There are also studies of antibiotic resistant *E. coli* and other organisms that are based on mass migration during religious festivals within Saudi Arabia. Specifically, Umrah and Hajj are two of

the most important religious pilgrimages in the world, and they both take place annually in the Kingdom of Saudi Arabia. Because of the high risk of infectious disease transmission associated with these occurrences, the Kingdom has the potential to become a hub for the dissemination of bacterial strains that are resistant to antimicrobials across the rest of the world. As part of the Arab Gulf region's attempt to lower the danger of AMR, the Saudi Ministry of Health (MOH) launched a nationwide antimicrobial stewardship plan (ASPs) in 2014 that included the adoption and implementation of ASPs in Saudi hospitals (Alghamdi et al. 2021).

4.1.2.1 ST131 in Saudi Arabia

Previously published research has shown that the most widespread sequence type among ExPEC STs found in Saudi Arabia is *E. coli* ST131 (Alghoribi et al. 2015; Alyamani et al. 2017; Yasir et al. 2018; Alqasim et al. 2020) (**Table 4.1**). The percentage of all ExPEC isolates in which ST131 was present ranged from 17.3 to 61.7% depending on the location (**Table 4.1**). The H30 clone was found to be the most frequent ST131 sub-clone, and H30 isolates were found to be more common than other ST131 sub-clones in terms of ESBL carriage and fluoroquinolone resistance as has been found in many other countries. However, because most of these studies are focussed on antibiotic resistant infections, they provide little information on the types of *E. coli* that are carried in the community or the actual normal carriage rates of resistant *E. coli*. As compared to other ExPEC clones, Alghoribi and others have demonstrated that this particular clone ST131 has a significantly higher association with ESBL production and antibiotic resistance which is characteristic of this *E. coli* ST wherever found (Alghoribi et al. 2015). Abd El Ghany and colleagues have analysed 10 carbapenem-resistant ExPEC isolates that were isolated from bacteraemia. During this process, they discovered the existence

of one *E. coli* ST131 strain that had the *bla*$_{NDM-1}$ gene for the New Delhi metallo β-lactamase (Abd El Ghany et al. 2018).

It has been revealed that the worldwide distribution of ESBLs, especially among ExPEC, is to account for the current high rates of resistance to β-lactam medicines (Alqasim et al. 2018). A prior study found that the prevalence of ESBL-production among ExPEC isolates in Saudi Arabia was 6.5% in the year 2002 (Kader et al. 2004). This finding is relevant to the question of how common ESBL-producing ExPEC are in the country. During the past few years, its prevalence has dramatically grown, going up to between 20.4% and 41.9% of all ExPEC isolates (Al-Agamy et al. 2014; Mously et al. 2016; Al Yousef et al. 2016; Alqasim et al. 2018). Genetic analysis of ESBL genotypes in ExPEC isolates revealed that *bla*$_{CTX-M}$ genes, which encode for CTX-M enzymes, were far more common than other β-lactamases expressing genes such as *bla*$_{TEM}$, *bla*$_{SHV}$, and *bla*$_{OXA}$, and that *bla*$_{CTX-M-15}$ was the most common CTX-M variant. It is interesting to note that the dissemination of the CTX-M-15 enzyme has been related with high levels of resistance among ExPEC isolates to penicillins and cephalosporins and also with co-carriage of other resistance mechanisms conferring resistance to amino-glycosides, nitrofurans, and fluoroquinolones (Al-Agamy et al. 2014; Alqasim et al. 2018).

Table 4.1. Previous studies showing the sequence types of *E. coli* in Saudi Arabia

City	Year of sample collection	Type of Sample	Number of *E. coli* isolates	Predominant STs	Other major STs	Reference
Riyadh	2012-2013	Urine	202	ST131	ST73, ST38, ST69, ST10, ST127, & ST95	Alghoribi *et al.*
Makkah	2014-2015	Urine	58	ST131	ST10, ST38, & ST648	Alyamani *et al.*
Jeddah	2014-2015	Urine, blood, and wound swab	211	ST131	ST38	Yasir *et al*.
Riyadh	2014-2015	Urine	10	ST448, & ST23	ST131 & ST69	Abd El Ghany *et al.*

4.1.3 Kazakhstan

Infections of the urinary tract are the primary factor in the Republic of Kazakhstan's upward trending pattern in the incidence of diseases affecting the urinary system (UTI). So, in 2015, the overall incidence in this nosological group was 7532.2, in 2016 the number rises up to 8784.8, and in 2017 to 8765.7 per 100,000 people (Muldaeva et al. 2019). In 2018, Kazakhstan was one of the nations in the European Area of the WHO that had a relatively low overall consumption of antibacterials. Despite this, the country did not fulfil the WHO national monitoring aim of having at least 60% of total antibacterial consumption coming from the "Access" category. This group has antibiotics that work against a wide range of common susceptible pathogens and have a lower chance of becoming resistant than antibiotics in the other group. There is still improper use of antibiotics, even though it has decreased over the past several years. This is mainly due to sales of antibiotics without prescriptions, overprescribing, and a lack of guidelines on the appropriate use of antibiotics at the national level (Zhussupova et al. 2021). Otherwise, there is almost no information on *E. coli* types that are commonly carried in the community in Kazakhstan.

4.2 Methodology rationale

4.2.1 Saudi Arabia and Kazakhstan *E. coli* isolates

E. coli isolates were selected from single samples of sewage collected from Unaziah wastewater treatment plant, Saudi Arabia (06/06/2019) and Karaganda wastewater treatment plant, Kazakhstan (17/04/2019). Isolates were selected randomly based on colour on brilliance UTI plates that either included no antibiotic, cefotaxime @ 10mg.L-1 or ciprofloxacin @ 1mg.L-1. To remove bias, we diluted sewage so that only 5-10 red colonies were found on each plate and picked all of them irrespective

of size or depth of colour. All species ID were confirmed by MALDI-TOF and isolates that were not *E. coli* were discarded.

Table 4.2. Date of collection and number of *E. coli* isolated from two nations: Saudi Arabia, and Kazakhstan

Sample	Date collection	Date isolation	Non-selection	CTX 10mg/L selection	CIP 1mg/L selection
Unaizah, KSA	06/06/2019	18/02/2019	45	39	13
Karaganda, KA	17/04/2019	22/04/2019	43	56	16

The isolates were also assayed for antibiotic resistance by disc diffusion using CLSI guidelines. All isolates were phylotyped by the Doumith multiplex PCR (Doumith et al. 2015). All isolates selected without antibiotic selection and with selection on cefotaxime were Whole Genome Sequenced. We also sequenced about a dozen ciprofloxacin resistant isolates from each nation.

4.3 Results

4.3.1 Saudi Arabian *E. coli* phylotype, MLST and resistance genes

Analysis of the phylotypes of *E. coli* isolated without antibiotic selection informs us about the types of *E. coli* that are commonly carried in the community in these two different locations. The results indicate that the carriage of pathogenic B2 phylotype *E. coli* was very low in Unaizah, Saudi Arabia with only single examples of isolates belonging to B2 and D phylotypes and most isolates belonging to commensal phylotypes B1, C and A (**Figure 4.1**, **Table 4.4**). The MLST analysis indicated that the B2 isolate was of ST95, and the phylotype D isolate of ST3018. The most common commensal phylotypes included ST58, ST10, ST46 and ST401. To discover the *E. coli* types associated with common antibiotic resistance traits in Unaizah we also investigated the isolates that are commonly carried in this community and are resistant to cefotaxime and ciprofloxacin. Selection with cefotaxime caused a general shift in the phylotypes recovered (**Figure 4.2**& **4.7**, **Table 4.5**) with a near doubling in phylotype C (22%-41%) and reductions in phylotype A (31%-18%) and B1 (42%-26%) among the commensal phylotypes and with a large increase in phylotype D (2%-10%) and no change in phylotype B2 (3%-5%) among the pathogenic phylotypes (**Table 4.3**).

Table 4.3. Changes to common phylotypes with antibiotic selection in Saudi Arabia

Phylotype	No antibiotic selection %	Cefotaxime selection 10mg.L-1 %
A	31	18
B1	42	26
B2	3	5
C	22	41
D	2	10

Analysis of the MLST data indicated the large increase in phylotype C was mostly due to the isolation of cefotaxime resistant ST10 *E. coli.* Apart from ST10 the selection caused a drastic change in the most commonly detected MLST types with the identification of ST69 and ST648 among phylotype D and the loss of ST95 with phylotype B2 and the loss of common commensal types ST58 and ST46 as well as several others (**Figures 4.1-4.2**, **Tables 4.4-4.5**). Surprisingly, ST131 was not found at all in any of the Saudi Arabian *E. coli* collections. The small number of isolates selected with ciprofloxacin did not allow a comparison but identified ST617 as an important ciprofloxacin resistant MLST in this region of Saudi Arabia.

Table 4.4. Saudi Arabian *E. coli* selected without antibiotics (phylotype, MLST and resistance gene data)

Isolate	Doumith phylotype PCR				PHY	Achtmann multilocus ST alleles							Sequence	*fimH*	*Antibiotic resistance genes*										
ID	gadE	chuA	yjaA	tspA		adk	fumC	gyrB	icd	mdh	purA	recA	Type		*β-lactamase*	*Sulphonamide, trimethoprim, tetracycline, aminoglycoside, chloramphenicol, fos,macrolide*									
SNO1	Y	N	N	Y	**B1**	6	4	3	16	9	8	6	**ST942**	39	*EC-13*				*tet(34)*						
SNO2	Y	N	Y	Y	**A**	52	54	46	48	35	40	38	**ST747**	0	*EC-8*		*dfrA14*		*tet(34)*						
SNO3	Y	N	Y	Y	**C**	10	11	4	8	8	8	2	**ST10**	54	*EC*				*tet(34)*						
SNO4	Y	N	N	Y	**B1**	6	19	4	18	11	13	6	**ST5539**	38	*EC-18*				*tet(34)*						
SNO5	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	27	*EC-18, TEM1*	*sul2*	*dfrA5*		*tet(34)*	*aph(3)-Ib*					
SNO6	Y	N	Y	N	**C**	10	11	135	8	8	8	2	**ST744**	54	*TEM1*	*sul1,2*	*dfrA17*	*qnrs13*	*tet(34,A)*	*aph(3)-Ia/b*	*aph(6)-Id*	*aadA5*	*catA1*	*mph*	*mef(b)*
SNO7	Y	N	N	Y	**B1**	6	19	3	16	11	8	6	**ST1727**	31	*EC-18*	*sul2*	*dfrA14*		*tet(34,A)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO8	Y	N	Y	N	**C**	10	7	4	12	8	18	6	**ST5359**	54	*EC-15*				*tet(34)*						
SNO9	Y	N	N	Y	**B1**	6	4	14	16	24	8	14	**ST155**	30	*EC-18*				*tet(34)*						
SNO10	Y	N	Y	N	**C**	10	11	4	12	8	8	2	**ST218**	0	*EC*		*dfrA14*		*tet(34)*						
SNO11	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	43	*TEM1*	*sul2*	*dfrA14*	*qnrS1*	*tet(34,A)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO12	Y	N	N	Y	**B1**	6	4	14	16	24	8	6	**ST949**	121	*EC-18, TEM1*	*sul2*			*tet(34,A)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO13	Y	N	N	Y	**B1**	6	19	3	16	9	8	6	**ST196**	87	*EC-18, TEM1*	*sul2*	*dfrA5*		*tet(34)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO14	Y	N	N	N	**A**	1	107	7	11	7	3	7	**ST401**	233	*EC*										
SNO15	Y	N	N	N	**A**	80	4	33	16	7	8	6	**ST939**	60	*EC-18*				*tet(34)*					*fosA7*	
SNO16	Y	Y	N	N	**D**	12	58	54	344	1	2	47	**ST3018**	24	*EC*				*tet(34)*						
SNO17	Y	N	N	Y	**B1**	6	6	5	9	9	8	2	**ST13**	32	*EC-18*				*tet(34)*						
SNO18	Y	N	N	N	**A**	1	107	7	11	7	3	7	**ST401**	233	*EC-18*				*tet(34)*						
SNO19	Y	N	Y	N	**C**	10	99	5	91	8	7	2	**ST361**	54	*EC*										
SNO20	Y	N	N	N	**A**	6	4	1	95	69	8	20	**ST399**	34	*EC*				*tet(34)*						
SNO21	Y	N	N	Y	**B1**	6	4	14	16	24	8	14	**ST155**	30	*EC-18*	*sul1*	*dfrA5*		*tet(34)*						
SNO22	Y	N	N	Y	**B1**	6	19	3	16	9	8	6	**ST196**	31	*EC-18*				*tet(34)*					*fosA7*	
SNO23	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	34	*EC-18, TEM1*	*sul2*	*dfrA5*		*tet(34)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO24	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	27	*EC*				*tet(34)*						
SNO25	Y	Y	Y	Y	**A**	6	11	4	8	8	8	2	**ST48**	0											
SNO26	Y	N	Y	N	**C**	8	7	1	8	8	8	6	**ST46**	54	*EC-15*				*tet(34)*						
SNO27	Y	N	N	N	**A**	8	7	1	8	8	8	6	**ST46**	34	*EC-15, OXA-1*		*dfrA14*		*tet(34,B)*			*aadA1*	*catA1*		
SNO28	Y	N	N	Y	**B1**	9	6	33	131	24	8	7	**ST641**	25	*EC-13*				*tet(34)*						
SNO29	Y	N	N	Y	**B1**	64	7	1	1	8	8	6	**ST398**	54	*EC-15*				*tet(34)*						
SNO30	Y	N	N	Y	**B1**	~747	19	33	26	11	8	6	**ST602***	86	*EC-18*				*tet(34)*						
SNO31	Y	N	N	Y	**B1**	6	6	5	16	11	8	7	**ST448**	31	*EC-13*			*qnrD1*							
SNO32	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24	*EC*				*tet(34)*						
SNO33	Y	N	N	N	**A**	8	7	1	8	8	8	6	**ST46**	34	*EC-15, TEM1, OXA1*	*sul2*			*tet(34,A)*	*aph(3)-Ib*	*aph(6)-Id*	*aadA1*	*catA1*		
SNO34	Y	N	Y	N	**C**	10	~ 7	5	8	8	8	2	**ST3856**	23	*EC-15*				*tet(34)*						
SNO35	Y	N	N	N	**B2**	37	38	19	37	17	11	26	**ST95**	15	*EC-5 -18*				*tet(34)*						
SNO36	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	30	*EC-18, TEM1*	*sul2*			*tet(34,A)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO37	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	27	*EC-18, TEM1*	*sul2*	*dfrA5*		*tet(34)*	*aph(3)-Ia/b*	*aph(6)-Id*				
SNO38	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	34	*EC-18, TEM1*				*tet(34)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO39	Y	N	Y	N	**A**	6	11	4	8	8	8	2	**ST48**	34	*EC-15, TEM1*	*sul3*	*dfrA14*	*qnrS13*	*tet(34,A)*		*aph(6)-Id*				*mef(B)*
SNO40	Y	N	Y	N	**C**	8	7	1	8	8	8	2	**ST1421**	34	*EC-15*				*tet(34)*						
SNO41	Y	N	N	N	**A**	113	4	5	83	8	8	6	**ST607**	305	*EC*	*sul1*			*tet(34)*						
SNO42	Y	N	N	Y	**B1**	6	19	3	18	9	8	6	**ST1056**	31	*EC-18*	*sul2*	*dfrA5*		*tet(34,A)*	*aph(3)-Ib*	*aph(6)-Id*				
SNO43	Y	N	N	N	**A**	6	107	1	95	69	8	7	**ST635**	34	*EC*				*tet(34)*						
SNO44	Y	N	N	N	**A**	6	4	1	95	69	8	20	**ST399**	34	*EC*				*tet(34)*						

SNO45	Y	N	N	Y	**B1**	1	107	7	11	7	3	7	**ST401**	233	***EC-18***				***tet(34)***						

Saudi Arabian *E. coli* selected without antibiotics (phylotype, MLST and resistance gene data). PHY indicates phylotype as designated by results of Doumith PCR. The results of the Doumith phylotype PCR are given as Y/N for presence or absence of a PCR product for each allele. ST were designated by in silico analysis of Achtman MLST alleles, individual allele numbers are given for each allele. An asterix for isolate SNO30 indicates that an exact ST could not be determined, and the closest ST (matching 6/7 alleles) is given. Resistance genes were determined from the in silico MISEQ sequence data using the CARD database. Isolate number =45.

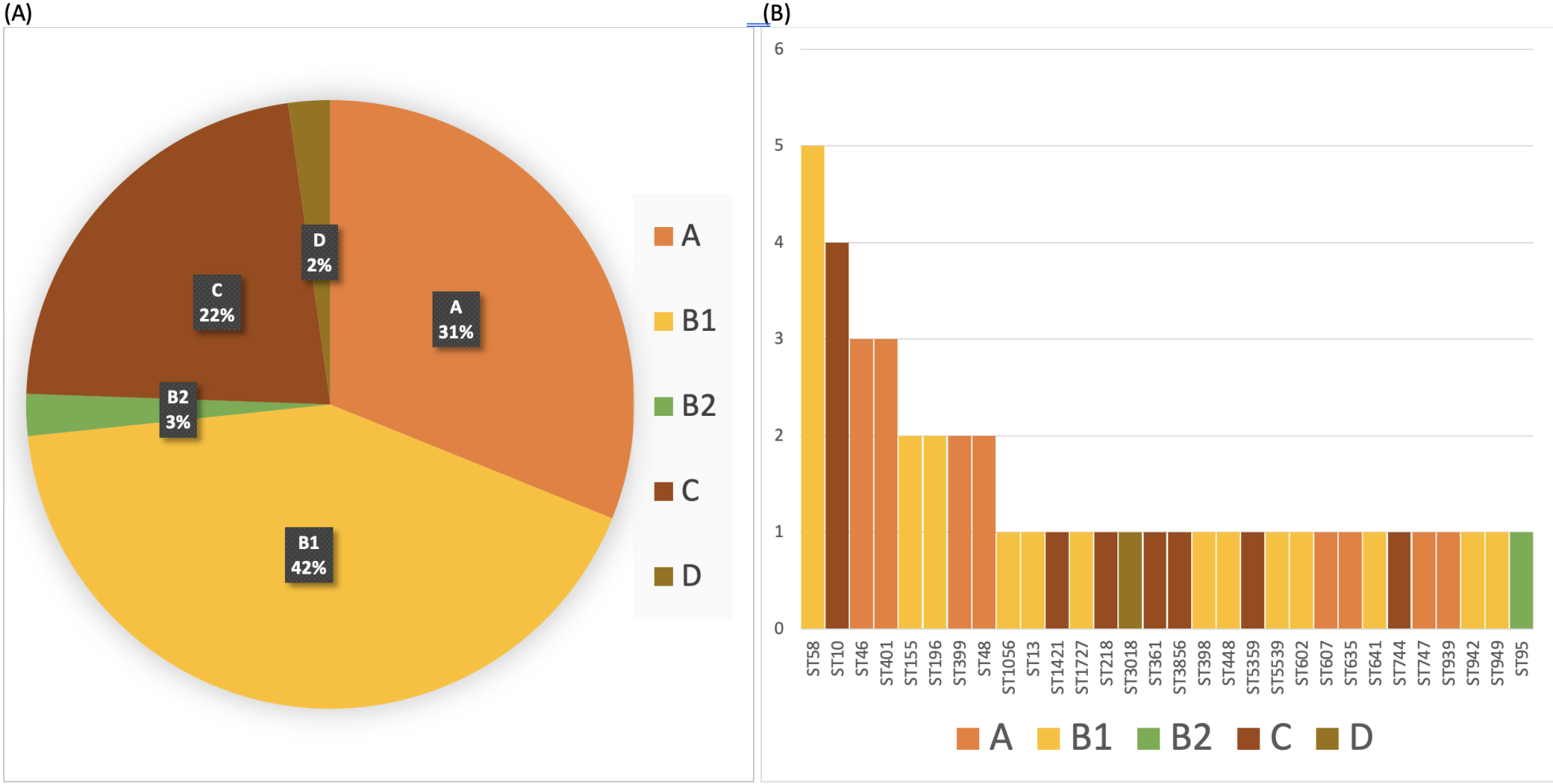


Figure 4.1. *E. coli* Phylotypes and MLST isolated from sewage in Saudi Arabia and isolated without antibiotic selection
(A) Pie chart indicates the proportions of *E. coli* phylotypes found in sewage from S. Arabia (Unaizah). (B) The bar chart indicates the most common multi-locus sequence types of *E. coli* that are found in sewage from S. Arabia (Unaizah). Total number of isolates, n=45.

Table 4.5. *E. coli* isolated from sewage in Saudi Arabia selected with cefotaxime 10mg.L-1

Isolate ID	Doumith phylotype PCR				PHY	Achtmann multilocus Alleles							Sequence Type	*fimH*	*Antibiotic resistance genes*												
	gadE	chu	yjA	tsp		adk	fumc	gyrB	icd	mdh	purA	recA			*β-lactamase genes*	*Tetracycline,quinolone.sulphnamide,aminoglycoside,colistin,macrolide/phenicol resistance genes*											
SCX1	Y	N	N	N	**A**	9	7	1	8	24	8	7	**ST6215**	34	***EC-15, CTX-M-14,OXA1***	***tet(34,A)***		***dfrA17***	***sul2***	***aadA5***	***aph(6)Id***	***aac(3)Id***		***mphA***		***floR***	
SCX2	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	28	***EC, CTX-M-15***	***tet(34)***	***qnrS1***										
SCX3	Y	N	N	Y	**B1**	6	4	33	16	11	8	6	**ST224**	61	***EC-18, CTX-M-15***	***tet(B)***											
SCX4	Y	N	N	Y	**B1**	57	1366	1	109	7	8	2	**ST515***	23	***EC-15, CTX-M-103***	***tet(34)***		***dfrA14***									
SCX5	Y	Y	N	Y	**D**	34	36	39	87	67	16	4	**ST349**	54	***EC-8, CTX-M-15***	***tet(34)***		***dfrA5***	***sul1***								
SCX7	Y	N	N	Y	**B1**	6	4	33	16	11	8	6	**ST224**	61	***EC-18, CTX-M-15, TEM1,***	***tet(A)***		***dfrA14***	***sul2,3***	***aadA1,2***	***aph(6)Id***	***aph(3)Iab***		***cmlA1***	***mefB***		
SCX8	Y	N	N	Y	**B1**	6	6	5	16	11	26	6	**ST5614**	54	***EC-18, CTX-M-15, TEM1,***	***tet(34)***											
SCX9	Y	N	Y	N	**C**	10	616	4	8	8	13	73	**ST4981**	0	***EC-15, CTX-M-15, TEM1,***	***tet(34)***	***qnrS1***	***dfrA14***	***sul2***		***aph(6)Id***	***aph(3)Ib***					
SCX10	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	215	***EC, CTX-M-15, TEM169, OXA1***	***tet(34,AB)***		***dfrA14***	***sul3***	***aadA1,2***	***aph(6)Id***	***aph(3)Iab***	***mcr-1***	***cmlA1***	***mefB***	***floR***	***aph(4)-Ia***
SCX11	Y	Y	N	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC-8, CTX-M-15***	***tet(A)***	***qnrS1***	***dfrA14***	***sul2***		***aph(6)Id***	***aph(3)Ib***					
SCX13	Y	N	N	N	**A**	8	7	4	8	8	512	6	**ST7588**	0	***EC, CTX-M-15, TEM1,***	***tet(34,A)***	***qnrS1***	***dfrA1***	***sul1,2,3***	***aadA1,2***	***aph(6)Id***	***aph(3)Ib***		***cmlA1***	***mefB***		
SCX14	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24	***EC, CTX-M-15***	***tet(34)***	***qnrS1***	***dfeA17***	***sul1***	***aadA5***				***mphA***			
SCX15	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC, CTX-M-15***	***tet(34,A)***		***dfrA1***		***aadA1***							
SCX16	Y	Y	Y	Y	**B2**	13	13	9	13	16	5	9	**ST4363**	0	***EC-5, CTX-M-15, TEM1,***				***sul2***		***aph(6)Id***	***aph(3)Ib***					
SCX17	Y	N	N	Y	**B1**	6	4	33	16	11	8	6	**ST224**	61	***EC-18, CTX-M-15, TEM1,***	***tet(A)***		***dfrA14***	***sul2,3***	***aadA1***	***aph(6)Id***	***aph(3)Iab***		***cmlA1***	***mefB***		
SCX18	Y	N	N	N	**A**	8	7	4	8	8	512	6	**ST7588**	0	***EC, CTX-M-15, TEM1,***	***tet(34,A)***	***qnrS1***	***dfrA14***	***sul2***		***aph(6)Id***	***aph(3)Ib***					
SCX19	Y	N	Y	N	**C**	6	4	12	1	20	12	7	**ST88**	171	***EC-13, TEM1,***	***tet(B)***	***qnrS1***	***dfraA1***	***sul2***	***aadA1***	***aph(6)Id***	***aph(3)Ib***					
SCX21	Y	N	Y	N	**C**	10	11	4	8	8	13	73	**ST617**	0	***EC-15, TEM33, OXA1***	***tet(34)***					***aph(6)Id***	***aph(3)Ib***					
SCX22	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24	***EC, CTX-M-15***	***tet(34)***	***qnrS1***	***dfrA17***	***sul1***	***aadA5***				***mphA***			
SCX23	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	215	***EC, CTX-M-15, TEM1, OXA1***	***tet(34,AB)***		***dfrA14***	***sul3***	***aadA1,2***	***aph(6)Id***	***aph(3)Ib***	***mcr-1***	***cmlA1***		***floR***	***aac(6')-Ib-cr***
SCX24	Y	N	N	Y	**B1**	6	4	33	16	11	8	6	**ST224**	61	***EC-18, TEM1,***												
SCX27	Y	Y	N	N	**D**	92	580	87	96	70	58	2	**ST4553**	32	***EC-19, CTX-M-15***	***tet(34)***	***qnrS1***										
SCX28	Y	Y	N	Y	**D**	92	4	87	96	70	58	2	**ST648**	0	***EC-19, CTX-M-15***	***tet(34)***	***qnrS1***										
SCX29	Y	N	N	N	**A**	8	7	1	8	7	7	335	**ST11226**	0	***EC-15,***	***tet(34)***											
SCX30	Y	N	Y	N	**C**	10	616	4	8	8	13	73	**ST4981**	0	***EC-15, CTX-M-15 TEM1,***	***tet(34,A)***	***qnrS1***	***dfrA14***	***sul2***		***aph(6)Id***	***aph(3)Ib***					
SCX32	Y	N	N	Y	**B1**	6	6	5	16	11	8	7	**ST448**	31	***EC-13, CTX-M-15 TEM1,***		***qnrS1***	***dfrA14***	***sul2***								
SCX34	Y	N	N	N	**A**	6	4	5	1	8	8	6	**ST4995**	27	***EC, CTX-M-15 TEM1,***	***tet(34)***											
SCX35	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24	***EC, CTX-M-10***	***tet(34)***	***qnrS1***	***dfrA17***	***sul1***	***aadA5***				***mphA***			
SCX36	Y	N	N	Y	**B1**	6	19	15	16	9	8	7	**ST443**	0	***EC-18, CTX-M-15, SHV-12, CMY145***	***tet(34)***	***qnrS1***										
SCX37	Y	N	N	N	**A**	6	7	57	1	8	8	2	**ST540**	54	***EC-15, CTX-M-15 TEM1,***	***tet(34)***	***qnrS1***		***sul2***			***aph(3)Ib***					
SCX38	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	28	***EC,***	***tet(34)***											
SCX39	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	28	***EC,***	***tet(34)***											
SCX40	Y	N	N	Y	**B1**	6	6	5	16	11	26	6	**ST5614**	54	***EC-18, CTX-M-15 TEM1,***	***tet(34,A)***	***qnrS1***	***dfrA17***	***sul1,2***	***aadA5***	***aph(6)Id***	***aph(3)Ib***		***mphA***			
SCX43	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	27	***EC, CTX-M-15, CMY-4***	***tet(34)***	***qnrS1***	***dfrA17***	***sul1,2***	***aadA5***				***mphA***			***erm(B)***
SCX44	Y	Y	Y	N	**B2**	58	53	53	58	24	1	42	**ST219**	370	***EC,***		***qnrS1***		***sul2***								
SCX46	Y	N	N	Y	**B1**	6	6	5	16	11	26	6	**ST5614**	54	***EC-18, CTX-M-15 TEM1,***	***tet(34)***	***qnrS1***										
SCX47	Y	N	N	N	**A**	64	196	5	83	8	8	6	**ST1972**	305	***EC, CTX-M-1***	***tet(34)***			***sul2***								
SCX48	Y	N	Y	N	**C**	10	11	4	8		8	2	**ST10**	54	***EC, CTX-M-15 TEM1, CMY-107***	***tet(34,A)***	***qnrS1***	***dfrA14***	***sul2***		***aph(6)Id***	***aph(3)Ib***					
SCX49	Y	N	Y	N	**C**	6	11	4	8	8	78	2	**ST1312**	198	***EC-15, CTX-M-15***	***tet(34)***	***qnrS1***	***dfrA14***	***sul2***		***aph(6)Id***	***aph(3)Ib***					

Saudi Arabian *E. coli* selected with cefotaxime 10mg.L^{-1} (phylotype, MLST and resistance gene data). PHY indicates phylotype as designated by results of Doumith PCR. The results of the Doumith phylotype PCR are given as Y/N for the presence or absence of a PCR product for each allele. ST were designated by in silico analysis of Achtman MLST alleles, individual allele numbers are given for each allele. An asterix for isolate SCX4 indicates that an exact ST could not be determined and the closest ST (matching 6/7 alleles) is given. Resistance genes were determined from the in silico MISEQ sequence data using the CARD database. Isolate number =39.

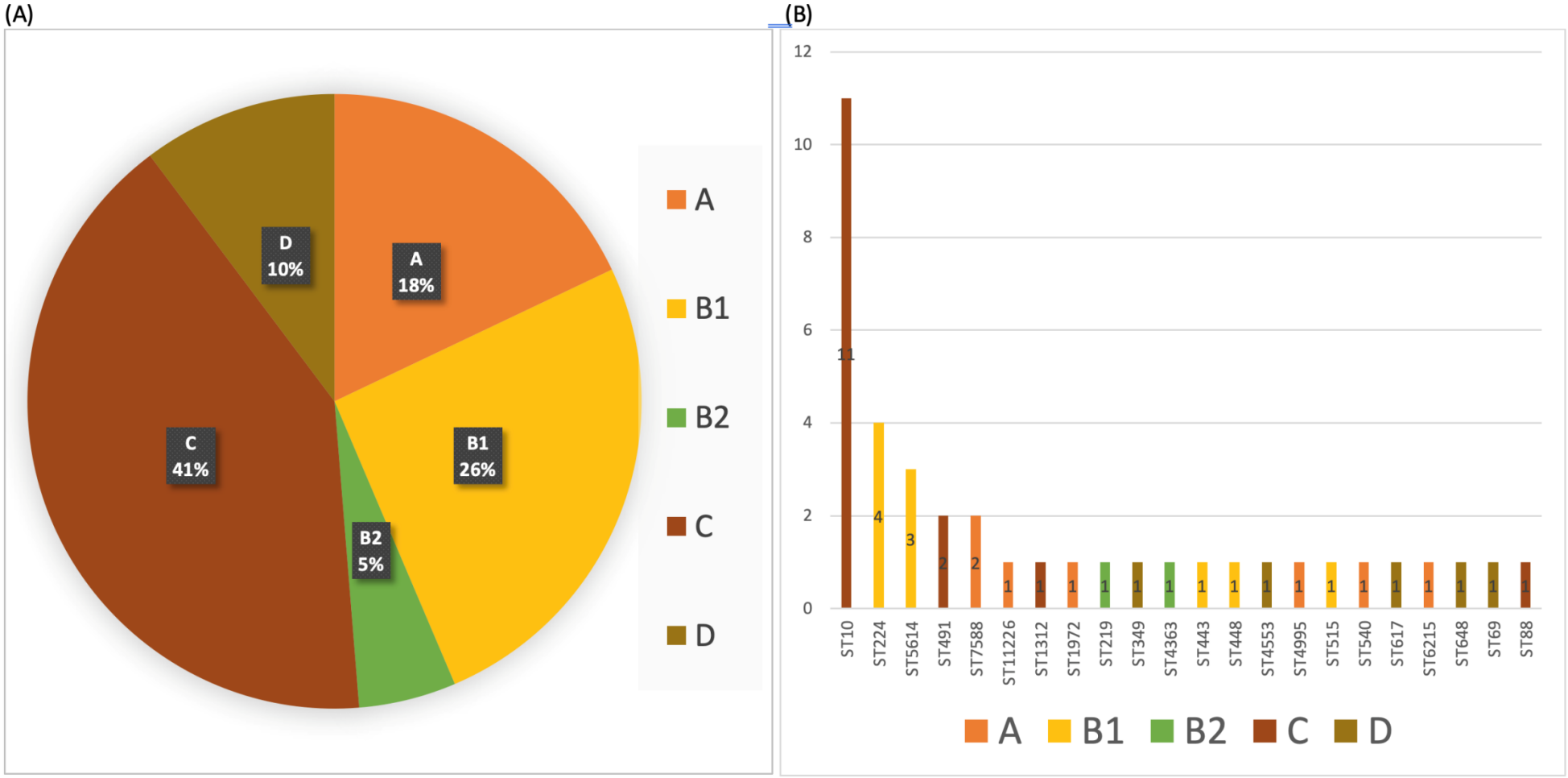


Figure 4.2. *E. coli* Phylotypes and MLST isolated from sewage in Saudi Arabia and selected with cefotaxime 10mg.L-1

(A) Pie chart indicates the proportions of cefotaxime resistant *E. coli* phylotypes found in sewage from S. Arabia (Unaziah). (B) The bar chart indicates the most commonly found multi-locus sequence types of *E. coli* that are resistant to cefotaxime found in sewage from S. Arabia (Unaziah). Total number of isolates, n=39.

Table 4.6. *E. coli* from Saudi Arabia isolated with ciprofloxacin 1mg.L-1

Isolate	Doumith phylotype PCR				PHY	Achtmann multilocus alleles							ST	*fimH*	Antibiotic resistance genes												
ID	gadE	chuA	yjaA	tspA		adk	fumC	gyrB	icd	mdh	purA	recA			β-lactamase genes	tetracyclin^{R}, aminoglycoside^{R}, sulphonamide^{R}, trimethoprim^{R}, chloramphenicol^{R}, macrolide^{R}, other												
SCP1	Y	N	N	Y	B1	80	95	3	18	11	8	333	ST5869	31	***EC18,***												
SCP2	Y	N	Y	N	C	10	11	4	8	8	13	73	ST617	0	***EC15, CTX-M15, OXA1, TEM1***	***tet(34,AB)***	***aph(6)Id***	***aph(3)Ib***		***aac(6')-Ib-cr***		***sul2***	***dfrA5***	***catB3***			
SCP4	Y	N	Y	N	C	6	11	4	10	7	8	6	ST93	30	***EC15***	***tet(34)***											
SCP6	Y	N	Y	N	C	10	11	4	8	8	13	73	ST617	0	***EC15, CTX-M15, OXA1***	***tet(34,B)***	***aph(6)Id***	***aph(4)Ia***	***aph(3)-Ib***	***aac(3)-IVa***	***aadA22***	***sul2***		***catB3***			
SCP8	Y	N	Y	N	C	10	11	4	8	8	13	73	ST617	27	***EC15, CTX-M15, OXA1***	***tet(34)***				***aac(6)-Ib-cr***				***catB3***			
SCP9	Y	N	N	Y	B1	6	65	32	26	11	8	6	ST8492	38	***EC18, TEM1***	***tet(A)***											
SCP10	Y	N	N	Y	B1	6	107	1	95	69	8	20	ST536	0	***EC15***	***tet(34)***											
SCP11	Y	N	Y	N	C	10	11	4	8	8	13	73	ST617	0	***EC15, CTX-M15, OXA1, TEM1***	***tet(34,AB)***	***aph(6)Id***	***aph(4)Ia***	***aph(3)-Ib***	***aac(6)-Ib-cr***	***aadA22***	***sul2***		***catB3***			
SCP12	Y	Y	N	N	D	4	26	2	25	5	5	19	ST38	5	***EC8, CTX-M15, OXA1, TEM1***	***tetA***	***aph(6)Id***		***aph(3)-Iab***		***aadA1***	***sul3***	***dfrA1***	***catB3***	***mefB***	***floR***
SCP15	Y	Y	N	N	D	34	36	39	87	67	16	4	ST349	54	***EC8, TEM1***	***tet(34,A)***						***sul1***	***dfrA5***	***catA1***			
SCP17	Y	Y	N	N	D	92	4	87	96	70	58	2	ST648	0	***EC19, CTX-M15***	***tet(34)***	***aph(6)Id***		***aph(3)-Ib***			***sul2***					
SCP26	Y	N	Y	N	C	10	11	4	8	8	8	2	ST10	54	***TEM1***	***tet(34)***											
SCP27	Y	N	N	Y	B1	43	41	15	18	11	7	6	ST101	86	***EC18***	***tet(34)***								***cmlA1***	***mphE***	***qnrS2***	

Saudi Arabian *E. coli* selected with ciprofloxacin 1mg.L^{-1} (phylotype, MLST and resistance gene data). PHY indicates phylotype as designated by results of Doumith PCR. The results of the Doumith phylotype PCR are given as Y/N for the presence or absence of a PCR product for each allele. ST were designated by in silico analysis of Achtman MLST alleles, individual allele numbers are given for each allele. Resistance genes were determined from the in silico MISEQ sequence data using the CARD database. Isolate number =13.

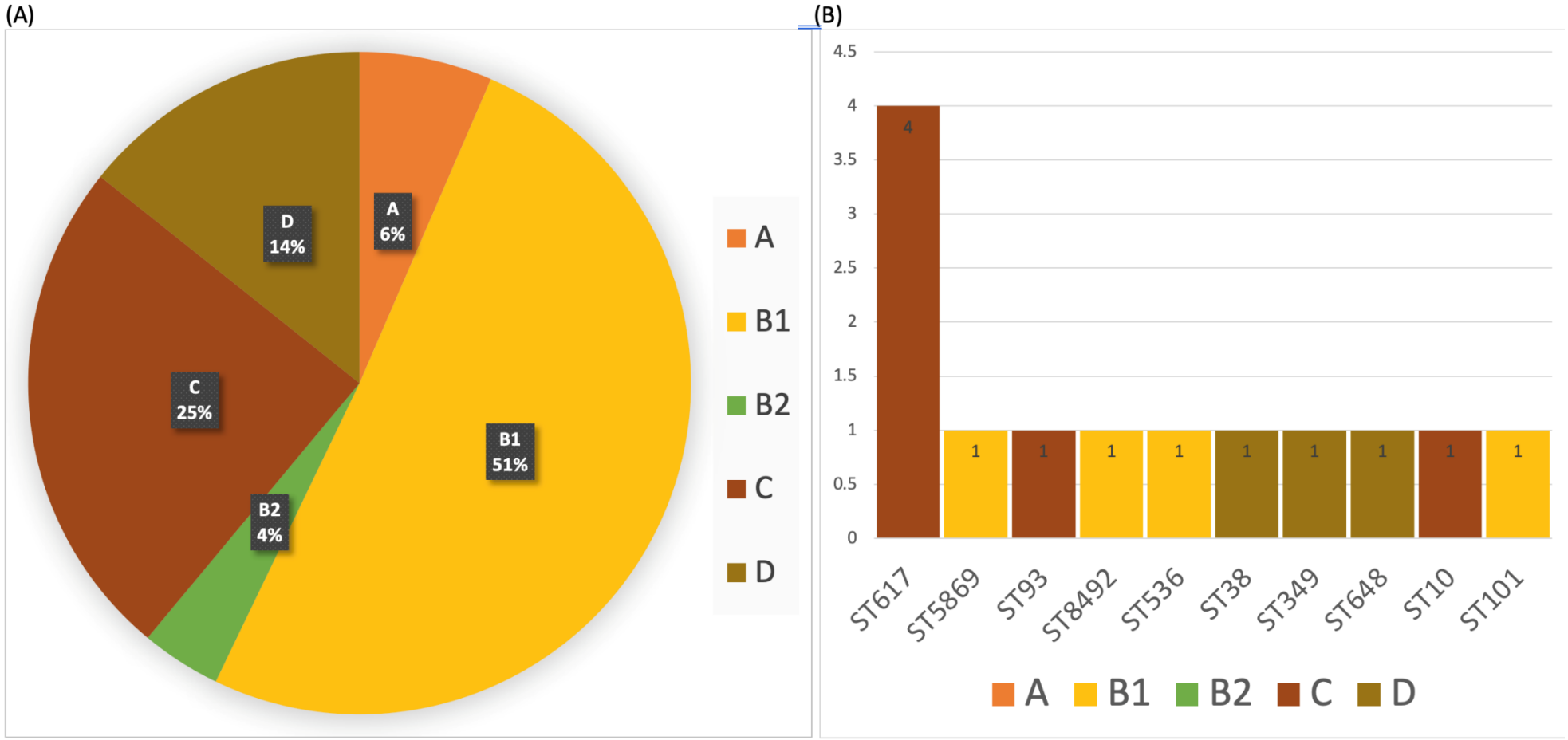


Figure 4.3. *E. coli* Phylotypes and MLST isolated from sewage in Saudi Arabia and selected with ciprofloxacin 1mg.L-1

(A) Pie chart indicates the proportions of *E. coli* phylotypes found in sewage from Saudi Arabia (Unaizah). Total number of isolates phylotyped n = 77. (B) The bar chart indicates 13 ciprofloxacin^{R} isolates that were randomly chosen for sequence analysis of resistance mechanisms and multi-locus sequence types of *E. coli* that are found in sewage from Saudi Arabia (Unaziah). Total number of isolates sequenced, n=13

4.3.2 Kazakhstan *E. coli* phylotype, MLST, and resistance genes

Analysis of the common phylotypes found in Kazakhstan revealed both similarities and differences to those found in Saudi Arabia. Phylotypes A and C were found in similar proportions in the different countries with phylotypes B1, B2 and D being quite different (**Table 4.7**). Generally, there were slightly more pathogenic phylotypes than commensal phylotypes in Kazakhstan relative to Saudi Arabia (9% B2+D in Kazakhstan verses 5% B2 + D in Saudi Arabia). However, in both countries *E. coli* isolated without antibiotic selection were overwhelmingly of commensal phylotypes (Saudi Arabia 95%, Kazakhstan 87% (**Table 4.7**).

Table 4.7. Comparison of *E. coli* phylotypes between Saudi Arabia and Kazakhstan

Phylotype	No selection		Cefotaxime selection (10mg.L-1)	
	Saudi Arabia %	Kazakhstan %	Saudi Arabia %	Kazakhstan %
A	31	41	18	5
B1	42	16	26	18
B2	3	9	5	25
C	22	30	41	22
D	2	4	10	30
Commensal (A+B1+C)	95	87	85	45
Pathogens (B2 +D)	5	13	15	55

On selection with cefotaxime the *E. coli* phylotype proportions changed drastically in Kazakhstan with a large reduction in phylotype A (41%-5%) and a smaller reduction in phylotype C (30%-22%) but with large increases in phylotypes B2 (9%-25%) and D (4% -30%). The overall change is a dramatic shift from commensal to pathogenic phylotypes (13%-55%) (**Table 4.7**).

Analysis of the MLST data revealed that there were similarities to the commensal MLST in that ST10 and ST58 were common MLST types in both countries (**Figures 4.1, 4.4**). However, whilst ST131 was absent from our Saudi Arabian *E. coli* isolated without antibiotic selection it was the most common pathogenic MLST in Kazakhstan

and other pathogenic MLST that are very common in Kazakhstan included ST95, ST394 and ST69, MLSTs that are often associated with bacteraemia. Analysis of the MLST associated with cefotaxime resistance revealed some interesting changes in commonly found MLSTs. Among commensal types ST10 was the most found commensal MLST Kazakhstan, similar to that found in Saudi Arabia. However, all other commonly found cefotaxime resistant MLSTs of *E. coli* belonged to pathogenic phylotypes B2 and D including ST131, ST69, ST394 and ST636, again these isolates are often found associated with bacteraemias. Ciprofloxacin selection also revealed that ST131 is an important MLST in Kazakhstan and also revealed other pathogenic MLST such as ST69 and ST38.

4.3.3 The genetic basis of antibiotic resistance in Saudi Arabia and Kazakhstan

Analysis of the Miseq data using the CARD database revealed some interesting information. Firstly, in all collections in both countries *tet34* was the most commonly found antibiotic resistance gene (**Tables 4.4-4.6** & **4.9-4.11**). The antibiotic resistance genes in *E. coli* without any antibiotic selection from Saudi Arabia were nearly double those of Kazakhstan, except for the *aadA* genes.

On selection with cefotaxime, the *E. coli* antibiotic resistance genes changed drastically between the two sites. Notably, several co-resistance genes were present with this selection that were not found in isolates without antibiotics including genes conferring macrolide, amphenicol and quinolone resistance, with the mobile quinolone resistance allele *qnrS* being particularly common in both nations. *aac(6'')-IB-cr* was commonly found only in Kazakhstan and it is known to not only confer resistance to aminoglycosides but also to confer reduced susceptibility to the commonly used fluoroquinolone antimicrobials, ciprofloxacin and norfloxacin (Robicsek et al. 2006).

Table 4.8. Comparison of *E. coli* antibiotic resistance genes between Saudi Arabia and Kazakhstan

Resistance gene	No selection		Cefotaxime selection (10mg.L-1)	
	Saudi Arabia %	Kazakhstan %	Saudi Arabia %	Kazakhstan %
sul	33	19	62	63
dfrA	30	16	33	54
aph(3")-Ib	28	16	44	25
aph(6)-Id	28	16	41	41
aadA	7	16	33	54
mph	-	-	15	20
cmlA	-	-	13	7
qnrS	-	-	54	59
cat	-	-	15	23
aac(6")-IB-cr	-	-	-	13
cmlA	-	-	13	7
mefB	-	-	10	4

Table 4.9. Kazakhstan *E. coli* isolated from sewage without antibiotic selection (phylotype, MLST and resistance gene data)

Isolate	Doumith phylotype PCR				PHY	Achtmann multilocus alleles							ST	*fimH*	Antibiotic resistance genes											
ID	*gadE*	*chuA*	*yjaA*	*tspA*		*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*			β-lactamase	tetracycline, quinolone, sulphonamide, trimethoprim, aminoglycoside, chloram/florphenicol, macrolide										
KNO1	Y	N	Y	N	**C**	10	11	4	1	8	9	2	**ST227**	94	***EC-15***	***tet(34,B)***										
KNO2	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	24	***EC-18***	***tet(34)***	***qnrB19***									
KNO3	Y	N	N	N	**A**	113		5	83	8	8	6	**ST607**	305	***EC***	***tet(34)***										
KNO4	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	137	***EC***	***tet(34)***										
KNO5	Y	N	N	N	**A**	6	107	1	95	69	8	7	**ST635**	34	***EC***	***tet(34)***										
KNO6	Y	N	N	N	**A**	6	107	1	95	69	8	7	**ST635**	34	***EC***	***Tet(34)***										
KNO7	Y	N	N	Y	**B1**	6	4	14	16	24	8	14	**ST155**	121	***EC-18***	***tet(34)***										
KNO8	Y	N	N	N	**A**	113	4	5	83	8	8	6	**ST607**	305	***EC***	***tet(34)***										
KNO9	Y	N	N	Y	**B1**	9	6	33	131	24	8	7	**ST641**	595	***EC-13***	***tet(34,A)***										
KNO10					**A**	6	107	1	95	69	8	7	**ST635**	34	***EC***	***tet(34)***										
KNO11	Y	N	N	N	**A**	6	107	1	95	69	8	7	**ST635**	34	***EC***	***tet(34)***										
KNO12	Y	N	N	N	**A**	6	107	1	95	69	8	7	**ST635**	1547	***EC***	***tet(34)***										
KNO14	Y	N	Y	N	**C**	10	11	4	12	8	8	2	**ST218**	30	***EC***	***tet(34)***										
KNO15	Y	N	N	Y	**B1**	9	65	5	1	9	13	6	**ST162**	32	***EC-18***	***tet(34)***										
KNO16	Y	N	N	N	**A**	6	4	1	1	69	8	20	**ST4774**	346	***EC***	***tet(34)***										
KNO17	Y	N	N	N	**A**	6	4	1	1	69	8	20	**ST4774**	346	***EC***	***tet(34)***										
KNO18	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24/1546	***EC***	***tet(34,A)***	***qnrS1***	***sul2***			***aph(6)-Id***	***aph(3'')-Ib***				
KNO19	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC***	***tet(34)***										
KNO20	Y	Y	Y	Y	**C**	10	11	4	8	8	8	2	**ST10**	27	***EC-15, TEM1, CTX-M15***	***tet(34)***	***qnrS1***									
KNO21	Y	N	Y	N	**C**	10	11	4	1	8	9	2	**ST227**	0	***EC-15***	***tet(34)***										
KNO22	Y	N	N	N	**A**	6	11	4	10	7	8	6	**ST93**	27	***EC-15, TEM1***	***tet(34,A)***	***qnrS1***	***sul2***	***dfrA1***	***aadA1***	***aph(6)-Id***	***aph(3'')-Ib***	***catA1***			
KNO23	Y	Y	Y	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC-8, TEM150***			***sul2***			***aph(6)-Id***	***aph(3'')-Ib***		***floR***		
KNO24	Y	Y	Y	Y	**B2**	37	38	19	37	17	11	26	**ST95**	30	***EC-5***	***tet(34)***										
KNO25	Y	N	Y	N	**C**	10	11	57	8	8	8	2	**ST1137**	54	***EC***	***tet(34)***										
KNO26	Y	N	N	Y	**B1**	119	150	12	16	24	8	6	**ST661**	38	***EC-18***	***tet(34)***										
KNO27	Y	Y	N	N	**A**	21	35	61	52	5	5	4	**ST394**	0	***EC-8, TEM1***				***dfrA1***	***aadA1***						
KNO28	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC***	***tet(34)***										
KNO29	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5***	***tet(A)***		***sul1,2***	***dfrA17***	***aadA5***	***aph(6)-Id***	***aph(3'')-Ib***				
KNO30	Y	Y	N	N	**A**	113	4	5	83	8	8	6	**ST607**	47	***EC-8, TEM2***	***tet(A)***		***sul1***	***dfrA1***	***aadA1***						
KNO31	Y	N	N	N	**A**	6	11	4	10	7	8	6	**ST93**	41	***EC-15, TEM1***	***tet(34,B)***		***sul1,2***	***dfrA17***	***aadA5***	***aph(6)-Id***	***aph(3'')-Ib***	***catA1***		***mph(A)***	***aac(3)-IId***
KNO32	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	27	***EC***	***tet(34)***										
KNO33	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC***	***tet(34)***										
KNO34	Y	N	N	Y	**A**	113	4	5	83	8	8	6	**ST607**	305	***EC***	***tet(34)***										
KNO35	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	54	***EC-18***	***tet(34)***										
KNO36	Y	N	Y	N	**A**	6	4	12	1	20	13	7	**ST23**	32	***EC-13***											
KNO37	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***TEM1***	***tet(34,B)***		***sul2***	***dfrA17***	***aadA5***	***aph(6)-Id***	***aph(3'')-Ib***				
KNO38	Y	Y	N	N	**D**	62	100	17	31	5	5	4	**ST362**	96	***EC-8, TEM2***	***tet(34,A)***										
KNO40	Y	N	N	N	**A**	6	296	4	10	7	8	6	**ST1838**	27	***EC-15***	***tet(34)***										
KNO41	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, TEM1, CTX-M15***	***tet(A)***		***sul1,2***	***dfrA17***	***aadA5***	***aph(6)-Id***	***aph(3'')-Ib***			***mph(A)***	
KNO42	Y	N	N	N	**A**	10	11	57	8	7	18	6	**ST216**	69	***EC-15***	***tet(34)***										
KNO43	Y	N	N	Y	**B1**	6	4	4	16	24	8	14	**ST58**	31	***EC-18***	***tet(34)***										
KNO44	Y	N	N	N	**A**	10	11	57	8	7	18	6	**ST216**	69	***EC-15***	***tet(34)***										
KNO45	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24/1546	***EC***	***tet(34)***										

PHY indicates phylotype as designated by results of Doumith PCR. The results of the Doumith phylotype PCR are given as Y/N for presence or absence of a PCR product for each allele. ST were designated by in silico analysis of Achtman MLST alleles, individual allele numbers are given for each allele. Resistance genes were determined from the in silico MISEQ sequence data using the CARD database.

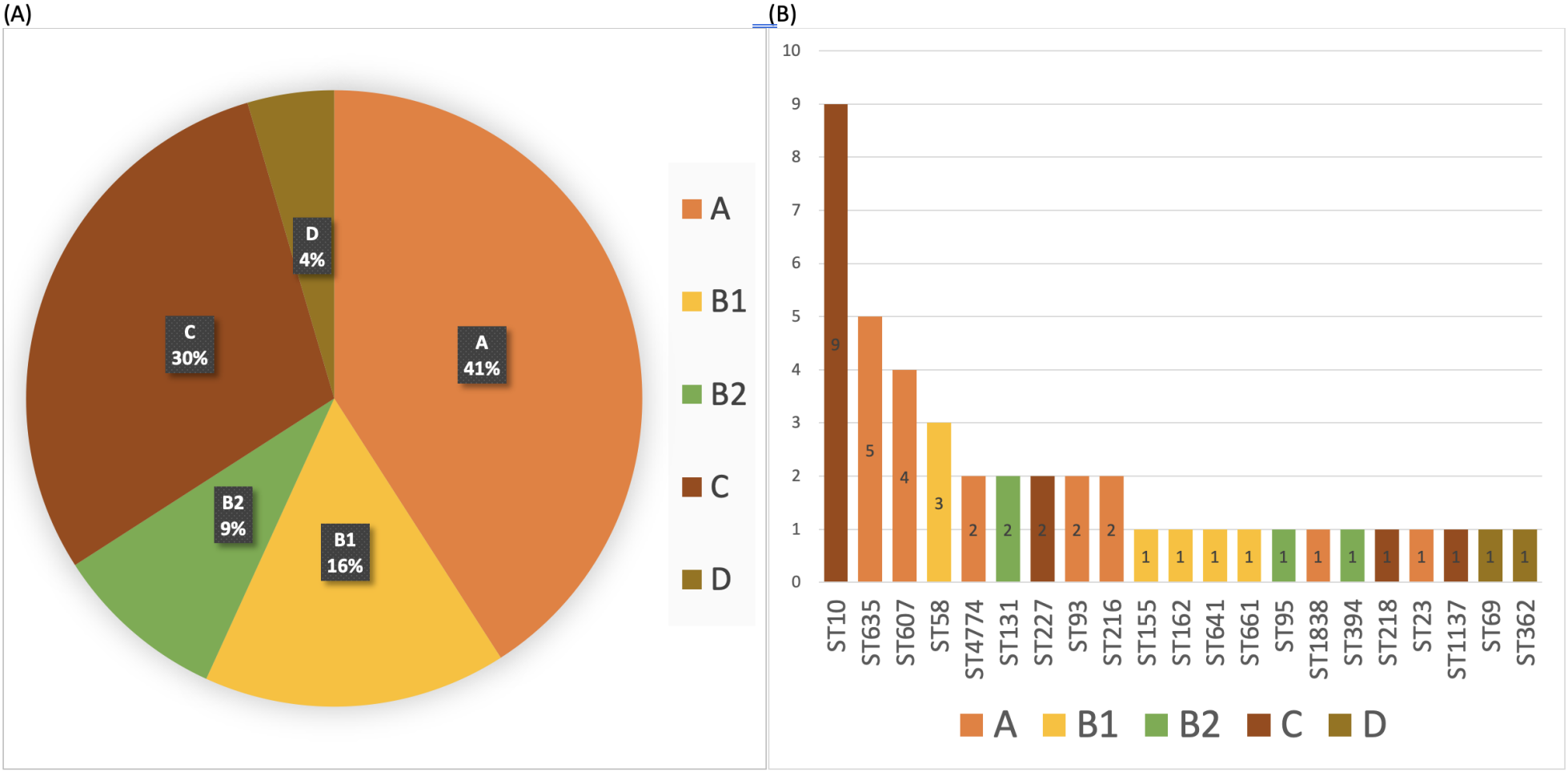


Figure 4.4. *E. coli* Phylotypes and MLST isolated from sewage in Kazakhstan without antibiotic selection

(A) Pie chart indicates the proportions of *E. coli* phylotypes found in sewage from Kazakhstan (Karaganda). (B) The bar chart indicates the most common multi-locus sequence types of *E. coli* that are found in sewage from Kazakhstan (Karaganda). Total number of isolates, n=43.

Table 4.10. Kazakhstan *E. coli* selected with cefotaxime 10mg.L-1

Isolate	*Doumith phylotype PCR*				PHY	Achtmann multilocus alleles							ST	*fim*	Antibiotic resistance genes										
ID	*gad*E	*chu*A	*yja*A	*tsp*A		*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*Pur*	*Rec*		H	β-lactamase genes	Tetracycline,quinolone,sulphonamide,trimethoprim, aminoglycoside, chloramphenicol,macrolide,other.									
KCX2	Y	Y	N	N	**D**	207	297	260	254	1	2	2	**ST2064**	0	***EC, CTX-M15, TEM1***	***tet(34,A)***	***qnrS1***	***sul1,2***	***dfrA1***	***aadA1***					
KCX3	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M27***										
KCX4	Y	Y	N	N	**D**	18	22	20	23	5	15	4	**ST414**	0	***EC-8, CTX-M14***	***tet(34)***									
KCX5	Y	N	N	N	**A**	6	107	~1	95	69	8	7	**ST635**	34	***EC, CTX-M15, TEM1, OXA1***	***tet(34,A)***	***qnrB1***	***sul2***	***dfrA14***	***aadA1***	***catA1,B3***	***aac(6')-Ib-cr***	***aac(3)-IIa***		
KCX8	Y	N	N	N	**A**	10	11	5	10	8	1	2	**ST1112**	41	***EC, CTX-M15***	***tet(34)***									
KCX9	Y	Y	Y	Y	**B2**	13	108	10	97	18	68	93	**ST636**	0	***EC-19, CTX-M15***				***dfrA1***	***aadA1***					
KCX10	Y	Y	Y	Y	**B2**									ND	***EC-19, CTX-M15***				***dfrA1***	***aadA1***					
KCX11	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC, CTX-M1, OXA1***	***tet(34,B)***	***qnrB19***	***sul1***			***catA1***				***floR***
KCX13	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15, OXA1***	***tet(A)***		***sul1***	***dfrA17***	***aadA5***	***catB3***	***aac(6'')-Ib-cr***	***aac(3)-IIa***	***mph(A)***	
KCX14	Y	N	Y	N	**C**	732	11	4	8	8	8	2	**ST7615**	54	***EC***	***tet(34)***	***qnrS1***								
KCX15	Y	Y	N	Y	**D**	~92	4	4	96	70	58	2	**ST648**	58	***EC-19, CTX-M14***	***tet(34,A)***		***sul1***		***aadA1***			***aac(3)-VIa***		
KCX17	Y	N	N	Y	**B1**	6	95	33	18	9	8	14	**ST442**	32	***EC-18***	***tet(34)***	***qnrS1***								
KCX18	Y	Y	Y	Y	**B2**	13	108	10	97	18	68	93	**ST636**	0	***EC-19, CTX-M15***				***dfrA1***	***aadA1***					
KCX19	Y	Y	Y	Y	**B2**	13	108	10	97	18	68	93	**ST636**	0	***EC-19, CTX-M15***	***tet(34)***			***dfrA1***	***aadA1***					
KCX20	Y	Y	N	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC-8, CTX-M15, TEM1***		***qnrS1***								
KCX21	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15, OXA1***	***tet(A)***		***sul1***	***dfrA17***	***aadA5***	***catB3***	***aac(6'')-IB-cr***	***aac(3)-Iia***	***mph(A)***	
KCX22	Y	N	N	Y	**B1**	9	65	5	1	9	13	6	**ST162**	54	***EC-13, CTX-M15, TEM1***	***tet(34)***	***qnrS1***	***sul2***	***dfrA1***	***aadA1***					
KCX23	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15, TEM1, OXA1***			***sul1***	***dfrA17***	***aadA5***	***catB3***	***aac(6')Ib-cr***		***mph(A)***	
KCX24	Y	N	N	Y	**B1**	6	29	32	16	11	8	44	**ST156**	38	***EC-18, CTX-M114, TEM1***	***tet(34)***	***qnrS1***	***sul1,2***	***dfrA1***	***aadA1***	***catA1***				
KCX26	Y	N	N	Y	**B1**	99	6	33	33	24	8	7	**ST453**	31	***EC-13, CTX-M169***	***tet(34)***		***sul2***	***dfrA17***	***aadA5***					***mcr-1***
KCX27	Y	Y	N	N	**D**	21	35	61	52	5	5	4	**ST394**	30	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul3***	***dfrA14***	***aadA1,2***	***cmlA1***			***mef(B)***	
KCX28	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	0	***EC, CTX-M15, TEM1***	***tet(34,A)***	***qnrS1***	***sul1,2***	***dfrA1***	***aadA1***	***catA1***				
KCX29	Y	Y	N	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul1,2***	***dfrA17***	***aadA5***			***aac(3)-IId***	***mph(A)***	
KCX30	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	27	***EC, CTX-M15, TEM1***	***tet(34,A)***	***qnrS1***	***sul1***	***dfrA17***	***aadA5***				***mph(A)***	
KCX31	Y	Y	N	N	**D**	20	45	41	43	5	32	~325	**ST117**	0	***EC, CTX-M14***	***tet(A)***		***sul1***	***dfrA1***	***aadA1***					
KCX32	Y	N	N	Y	**B1**	6	4	33	16	11	8	6	**ST224**	61	***EC-18, CTX-M15, TEM1, CMY2***	***tet(A)***	***qnrS1***	***sul1,2***	***dfrA14***	***aadA2***				***mph(A)***	
KCX33	Y	N	N	Y	**B1**	6	29	4	18	11	8	6	**ST212**	38	***EC-18, CTX-M55***										
KCX34	Y	Y	N	N	**D**	21	35	61	52	5	5	4	**ST394**	30	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul2,3***	***dfrA14***	***aadA2***	***cmlA1***				
KCX36	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15, OXA1***						***catB3***	***aac(6)-Ib-cr***	***aac(3)-Iia***		
KCX37	Y	Y	N	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul1,2***	***dfrA17***	***aadA5***		***aac(3)-Iia***	***aac(3)-Iid***	***mph(A)***	
KCX38	Y	Y	N	N	**D**	21	35	61	52	5	5	4	**ST394**	30	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul2,3***	***dfrA14***	***aadA1,2***	***cmlA1***			***mef(B)***	
KCX39	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	27	***EC, CTX-M101***	***tet(34,A)***	***qnrS1***								
KCX41	Y	Y	N	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC-8, CTX-M15, TEM1***		***qnrS1***								
KCX43	Y	Y	N	N	**D**	4	26	2	25	5	5	19	**ST38**	54	***EC-8, CTX-M15, TEM1***		***qnrS1***								
KCX44	Y	Y	N	N	**D**	92	4	87	96	70	58	2	**ST648**	58	***EC-19, CTX-M65***	***tet(34)***		***sul2,3***		***aadA1,2***	***cmlA1***			***floR***	***fosA3***
KCX46	Y	Y	N	N	**D**	21	35	61	52	5	5	4	**ST394**	30	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul2,3***	***dfrA14***	***aadA8***					
KCX48	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15, OXA1***	***tet(A)***					***catB3***	***aac(6')-Ib-cr***	***aac(3)-Iia***		***ac(3)-Iia***
KCX49	Y	Y	N	N	**C**	92	580	87	96	70	58	2	**ST4553**	32	***EC-19, CTX-M15, TEM1***	***tet(34)***	***qnrS1***	***sul2***	***dfrA14***						
KCX50	Y	Y	N	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC-15, CTX-M15, TEM1***	***tet(34)***	***qnrS1***								
KCX51	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M27***	***tet(A)***	***qnrB19***	***sul1,2***	***dfrA17***	***aadA5***					
KCX52	Y	Y	N	N	**D**	21	35	61	52	5	5	4	**ST394**	30	***EC-8, CTX-M15, TEM1***	***tet(A)***	***qnrS1***	***sul2***	***dfrA14***	***aadA2***	***cmlA1***			***mef(B)***	
KCX55	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	419	***EC, CTX-M15***	***tet(34)***								***mph(A)***	
KCX56	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	27	***EC, CTX-M15***	***tet(34)***	***qnrS1***	***sul1***	***dfrA17***	***aadA5***				***mph(A)***	
KCX57	Y	N	N	Y	**B1**	6	19	33	26	11	8	6	**ST602**	86	***EC-18, CTX-M1, CMY2,***	***tet(34,A)***		***sul1***	***dfrA1***	***aadA1***					***fosA7***

KCX58	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***CTX-M27***										
KCX59	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24	***EC, CTX-M15***	***tet(34,A)***	***qnrS1***	***sul2***	***dfrA14***						
KCX60	Y	Y	N	N	**D**	101	4	97	29	70	158	2	**ST1722**	153	***EC-19, CTX-M15***	***tet(34)***	***qnrS1***	***sul2***							
KCX61	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	24	***EC, CTX-M15***	***tet(34,A)***	***qnrS1***	***sul2***	***dfrA14***						
KCX62	Y	N	N	Y	**B1**	9	65	5	1	9	13	6	**ST162**	54	***EC-13, CTX-M15***	***tet(34)***	***qnrS1***	***sul2***							
KCX63	Y	N	Y	N	**C**	10	11	4	8	8	8	2	**ST10**	54	***EC, CTX-M139***	***tet(34,A)***	***qnrS1***								
KCX64	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15***	***tet(A)***		***sul1***							***mph(A)***
KCX65	Y	N	N	Y	**B1**	109	65	5	1	9	13	14	**ST517**	32	***EC-13, CTX-M15***	***tet(34)***	***qnrS1***	***sul1***	***dfrA17***	***aadA5***					***mph(A)***
KCX66	Y	N	N	Y	**B1**	9	65	5	1	9	13	6	**ST162**	54	***EC-13, CTX-M15***	***tet(34,A)***	***qnrS1***	***sul2***	***dfrA14***						
KCX68	Y	N	N	N	**A**	6	11	4	10	7	84	6	**ST484**	54	***EC-15, CTX-M15, TEM1***	***tet(34)***		***sul2***	***dfrA14***						
KCX69	Y	Y	N	N	**D**	92	580	87	96	70	58	2	**ST4553**	32	***EC-19, CTX-M15***	***tet(34)***	***qnrS1***								
KCX70	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC-5, CTX-M15, TEM1, OXA1***	***tet(A)***		***sul2***	***dfrA17***	***aadA5***	***catB3***	***aac(6')-Ib-cr***			

Kazakhstan *E. coli* selected with cefotaxime 10mg.L^{-1} (phylotype, MLST and resistance gene data). PHY indicates phylotype as designated by results of Doumith PCR. The results of the Doumith phylotype PCR are given as Y/N for presence or absence of a PCR product for each allele. ND indicates *fimH* not defined. ST were designated by in silico analysis of Achtman MLST alleles, individual allele numbers are given for each allele. Resistance genes were determined from the in silico MISEQ sequence data using the CARD database.

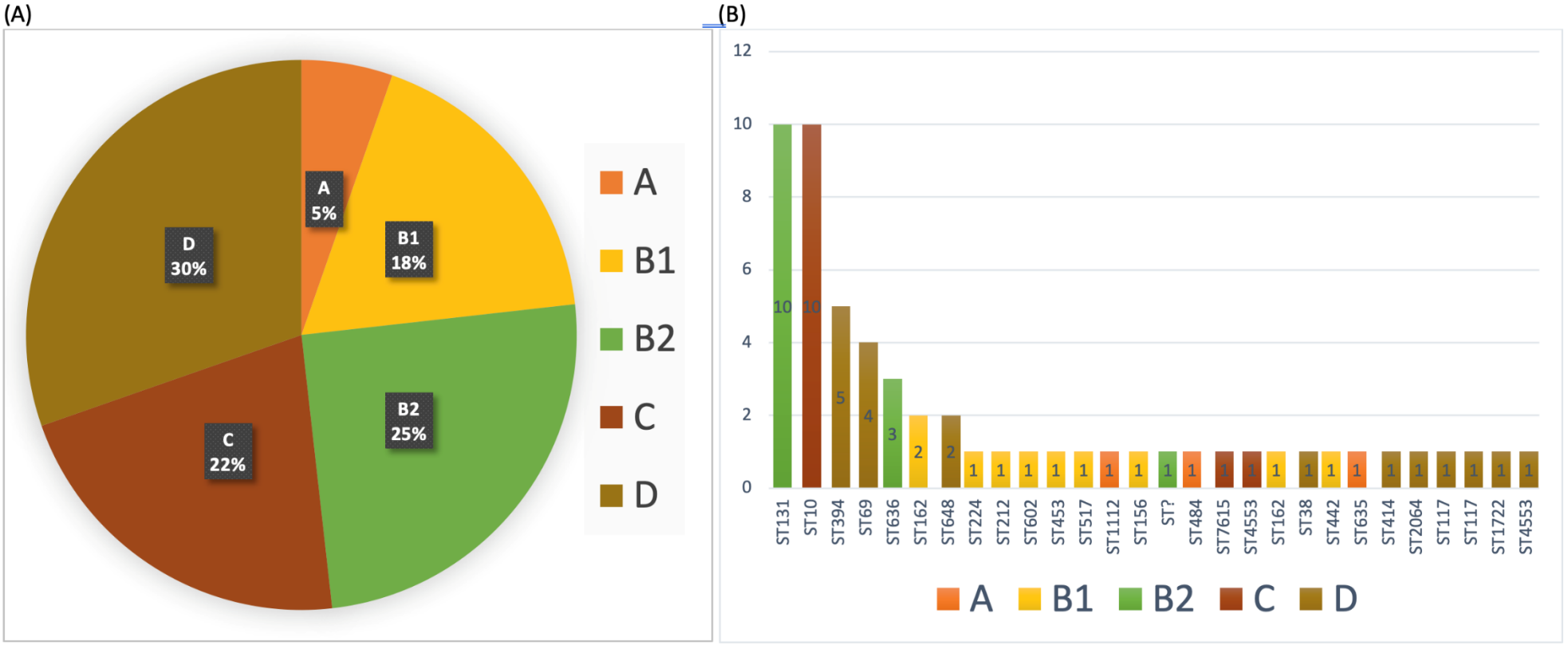


Figure 4.5. *E. coli* Phylotypes and MLST isolated from sewage in Kazakhstan and selected with cefotaxime 10mg.L-1

(A) Pie chart indicates the proportions of *E. coli* phylotypes found in sewage from Kazakhstan (Karaganda). (B) The bar chart indicates the most common multi-locus sequence types of *E. coli* that are found in sewage from Kazakhstan (Karaganda). Total number of isolates, n=56.

Table 4.11. Kazakhstan *E. coli* selected with ciprofloxacin 1mg.L-1

Isolate	Doumith phylotype PCR				PHY	Achtmann multilocus alleles							ST	*fimH*	Antibiotic resistance genes									
ID	*gadE*	*chu*	*yjiA*	*tsp*		*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*			β-lactamase^{R} genes	Aminoglycoside^{R}, trimethoprim^{R}, sulphonamide^{R}, chloramphenicol^{R}, macrolide^{R}, quinolone^{R}, fosphomicin^{R}									
KCP4	Y	N	Y	N	**C**	10	99	5	91	8	7	2	**ST361**	54	***EC, CTXM-15, CMY42, OXA1***		***aac(3)-IIa***	***aadA5***		***dfrA17***	***sul1***	***catA1,B3***	***mph(A)***		
KCP5	Y	N	Y	N	**C**	10	11	135	8	8	8	2	**ST744**	54	***EC, CTXM-15, TEM1, OXA1***	***aph(6)-Id***	***aph(3)-Ib***	***aadA1,5***		***dfrA1,17***	***sul2***	***catA1***	***mph(A,B)***	***qnrS1***	
KCP6	Y	N	Y	N	**C**	10	99	5	91	8	7	2	**ST361**	54	***EC, CTXM-15, CMY42, OXA1***		***aac(3)-IIa***	***aadA5***	***aac(6)Ib-cr***	***dfrA17***	***sul1***	***catA1,B3***	***mph(A)***		
KCP7	Y	Y	N	Y	**D**	4	26	2	25	5	5	19	**ST38**	65	***EC8, CTXM-1***		***aac(3)-IIa***	***aadA1***			***sul1***		***mphB***	***qnrB19***	
KCP9	Y	N	Y	N	**C**	10	174	4	8	8	8	2	**ST1286**	23	***EC15, TEM1***										
KCP10	Y	N	Y	N	**C**	10	11	135	8	8	8	2	**ST744**	54	***EC, CTXM-15, TEM1***	***aph(6)-Id***	***aph(3)-Ia***	***aadA1,5***		***dfrA1,17***	***sul1,2***	***catA1***	***mph(A,B)***		
KCP12	Y	N	N	Y	**B1**	6	19	33	26	11	8	6	**ST602**	86	***EC18, CTXM-1***	***aph(6)-Id***	***aph(3)-IIa,Ib***	***aadA1***		***dfra1***	***sul3***		***mefB***		***fosA7***
KCP13	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC5, CTXM-15, OXA1***		***aac(3)-IIa***	***aadA5***	***aac(6)Ib-cr***		***sul1***	***catB3***		***qnrS1***	
KCP16	Y	Y	N	N	**D**	21	35	27	6	5	5	4	**ST69**	27	***EC8, CTXM-14***		***aac(3)-Id***	***aadA1,2***		***aadA12***	***sul3***	***catA1***			
KCP18	Y	N	Y	N	**C**	10	11	135	8	8	8	2	**ST744**	54	***EC, CTXM-15, TEM1***	***aph(6)-Id***	***aph(3)-Ib***	***aadA5***		***dfrA17***	***sul2***	***catA1***	***mph(A)***	***qnrB19***	
KCP19	Y	N	N	Y	**B1**	9	65	5	1	9	13	6	**ST162**	32	***EC18, TEM1***										
KCP21	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC5, CTXM-15, OXA1***		***aac(3)-Ia***	***aadA5***		***dfrA17***	***sul1***	***catB3***	***mph(A)***		
KCP35	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC5, CTXM-101, TEM1***	***aph(6)-Id***	***aph(3)-Ib***	***aadA5***		***dfrA17***	***sul1,2***				
KCP37	Y	Y	N	Y	**D**	6	4	159	44	112	1	17	**ST1011**	31	***EC15, TEM1***	***aph(6)-Id***	***aac(3)-IId***	***aadA1***		***dfrA1,12***	***sul1***	***catA1***	***mph(A)***		
KCP48	Y	N	Y	N	**C**	10	99	5	91	8	7	2	**ST361**	54	***EC, CTXM-15, CMY42, OXA1***		***aac(3)-IIa***	***aadA5***	***aac(6)Ib-cr***	***dfrA17***	***sul1***	***catA1***	***mph(A)***		
KCP58	Y	Y	Y	Y	**B2**	53	40	47	13	36	28	29	**ST131**	30	***EC5, TEM1***	***aph(6)-Id***	***aph(3)-Ib***				***sul2***				

Kazakhstan *E. coli* selected with ciprofloxacin 1mg.L^{-1} (phylotype, MLST and resistance gene data). PHY indicates phylotype as designated by results of Doumith PCR. The results of the Doumith phylotype PCR are given as Y/N for presence or absence of a PCR product for each allele. ST were designated by in silico analysis of Achtman MLST alleles, individual allele numbers are given for each allele. Resistance genes were determined from the in silico MISEQ sequence data using the CARD database.

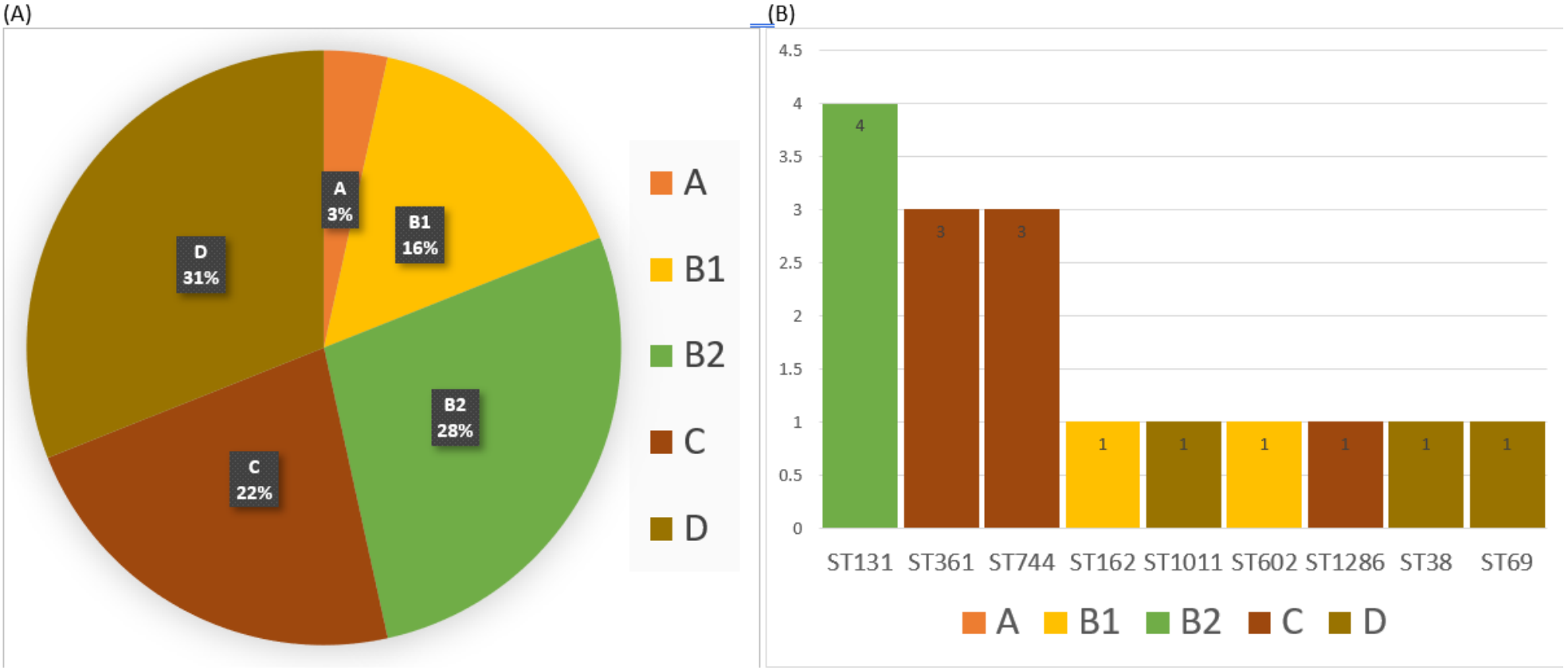


Figure 4.6. *E. coli* Phylotypes and MLST isolated from sewage in Kazakhstan and selected with ciprofloxacin 1mg.L-1

(A) Pie chart indicates the proportions of *E. coli* phylotypes found in sewage from Kazakhstan (Karaganda). Total number of isolates phylotyped n = 58. (B) The bar chart indicates 16 ciprofloxacin^{R} isolates that were randomly chosen for sequence analysis of resistance mechanisms and multi-locus sequence types of *E. coli* that are found in sewage from Kazakhstan (Karaganda). Total number of isolates sequenced, n=16

4.3.4 Antibiotic resistance among *E. coli* in Saudi Arabia and Kazakhstan

Antibiotic disc diffusion was performed on isolates from the two countries using 14 different antibiotics and antibiotic/inhibitor combinations: ciprofloxacin (CIP, 1µg), sulphonamides: sulfamethoxazole/trimethoprim (SXT, 1,25/ 23,75µg), aminoglycosides: gentamicin (CN, 30µg), β-lactams: amoxicillin (AML, 20µg), cefotaxime (CTX, 30µg) ceftazidime (CAZ, 30µg) tazobactam/piperacillin (TZP, 10/75µg), amoxycillin/clavulanic acid (AMC, 20/10µg), chloramphenicol (CL, 30µg), mecillinam (MIL, 10µg), fosfomycin (FOT, 200µg), nitrofurantoin (F, 100µg), trimethoprim (TM, 5µg), Amikacin (AK, µg). However, isolates with antibiotics selection were only tested for 8 antibiotics. As expected, the overall resistance for Saudi Arabia isolates without any antibiotic selection was overall low for and absent for several antibiotics e.g., amikacin and nitrofurantoin. The resistance for amoxicillin was 40% for A and B1 phylogroups. Gentamicin and trimethoprim resistance was higher in group D isolates. Isolates with no antibiotic selection from Kazakhstan were particularly high for group D in resistance to amoxicillin, gentamicin, and trimethoprim.

We determined the antibiotic resistance of isolates selected with cefotaxime and ciprofloxacin using 8 antibiotics. As expected, Saudi Arabian isolates selected with cefotaxime 10mg.L-^{1} show high resistance for amoxicillin and ceftazidime but low resistance for piperacillin/tazobactam and amoxiclav and show no resistance for amikacin. Kazakhstan isolates selected using cefotaxime 10mg.L^{-1} also showed low resistance to piperacillin/tazobactam. Saudi Arabian isolates selected using ciprofloxacin 1mg.L-^{1} showed high co-resistance for amoxicillin for all phylogroups but low resistance for other antibiotics. Kazakhstan isolates selected using ciprofloxacin 1mg.L-^{1} also showed high resistance for amoxicillin for all groups.

Phylogroup B2 showed high co-resistance to cefotaxime and ceftazidime but low resistance for other antibiotics (**Figure 4.7 and 4.8**).

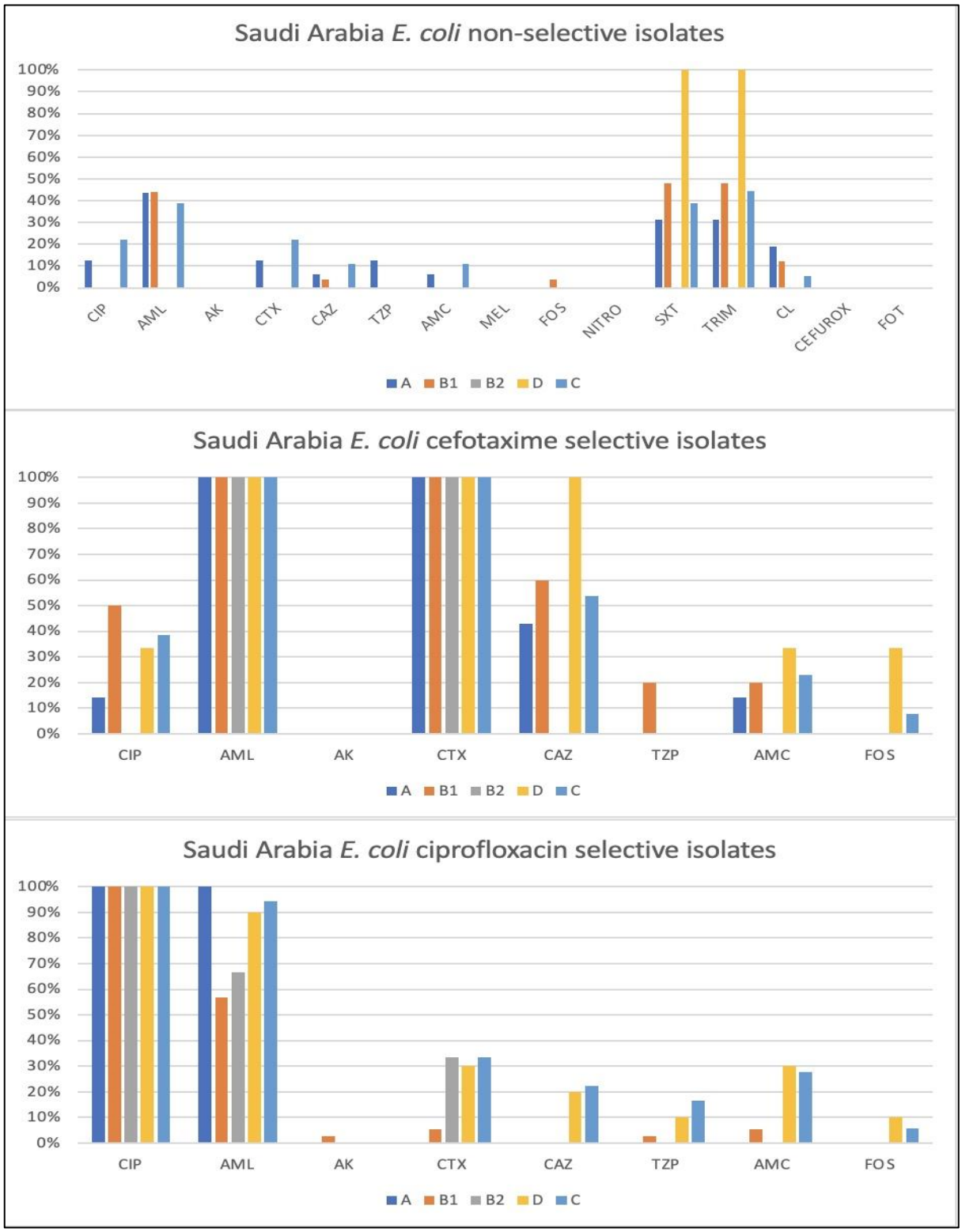


Figure 4.7. Percent resistance by phylotype for *E. coli* collected from Unaizah, Saudi Arabia for isolates without any antibiotic, with cefotaxime, or ciprofloxacin

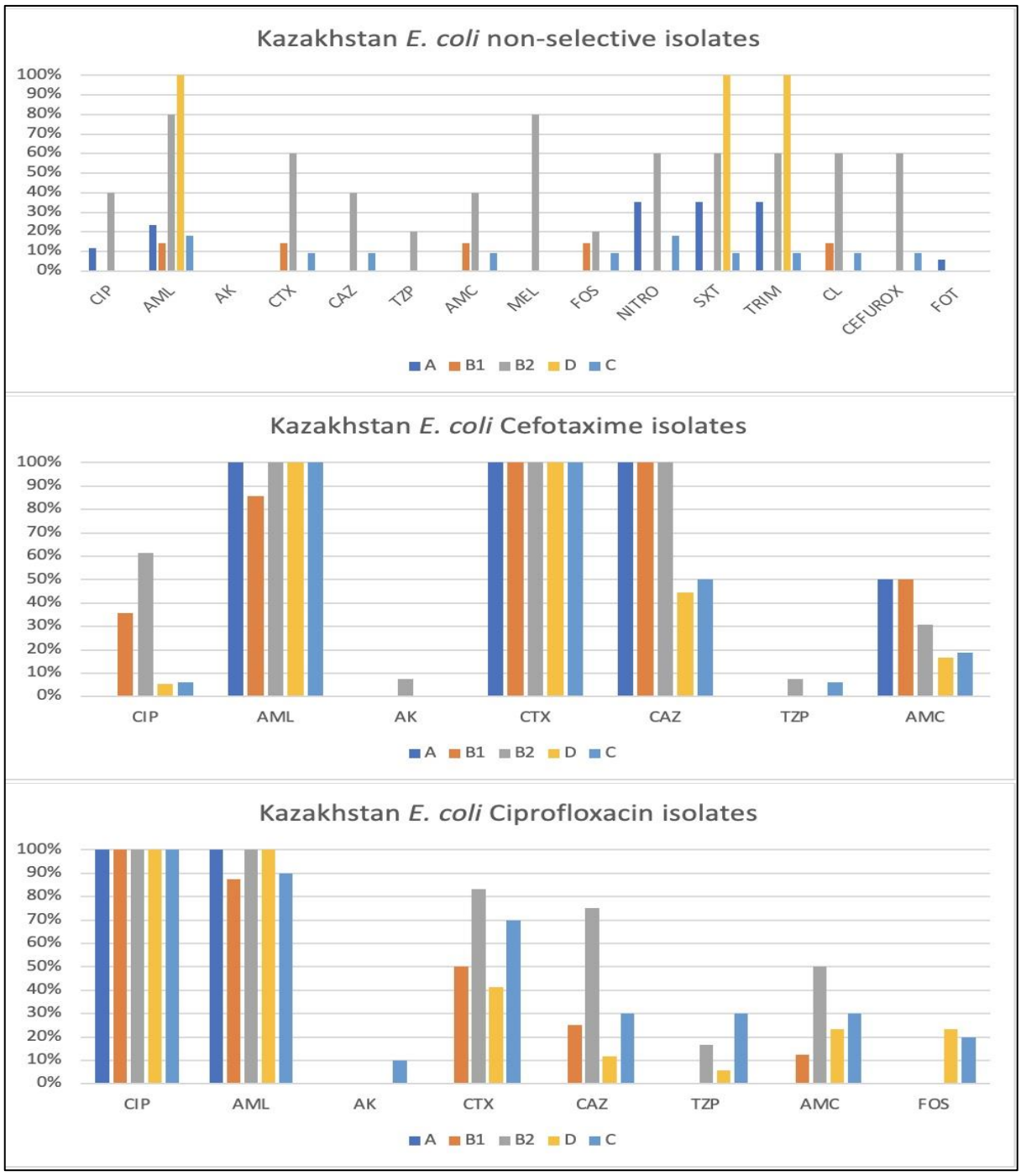


Figure 4.8. Percent resistance by phylotype for *E. coli* collected from Karaganda, Kazakhstan for isolates without any antibiotic, with cefotaxime, or ciprofloxacin

4.3.5 Minion long read sequencing and assemblies

Fifty-two isolates were sequenced using a single Nanopore minion flow cell FAS49777. Each DNA was individually barcoded using the SQK-RBK 110.96 barcoding kit before adding the pooled and barcoded DNA to the flow cell. The Minion flow cell was run for 72 hrs and barcoded FAST5 files were loaded onto the Hawk supercomputer in Cardiff and translated into FASTQ files using Guppy software. The FASTQ files were then assembled using Flye software and annotated using Prokka. Details of isolates and number of reads and assembly statistics are given in (**Table 4.12**) below. Assemblies were of various qualities which was mostly due to the depth of coverage. Most isolates were completely closed, and chromosomal/plasmid location was identified using local blast searches using Geneious software (Biomatters. USA).

Table 4.12. Nanopore minion flow cell/isolate ID and assembly statistics

NANOPORE run	Isolate/ Barcode	#FAST5 folders	#FASTQ folders	Miseq	Flye assembly (#contigs)	Length	COV
FAS49777	KCP9/BC1	13	13	Y	Y (2)	4,966,966	57 X
FAS49777	KCP13/BC2	23	23	Y	Y (2)	5,278,712	112 X
FAS49777	KCP18/BC3	21	21	Y	Y (2)	4,877,998	126 X
FAS49777	KCP21/BC4	24	24	Y	Y (2)	5,286,564	85 X
FAS49777	KCP35/BC5	10	10	Y	Y (4)	5,297,445	38 X
FAS49777	KCP37/BC6	13	13	Y	Y (4)	5,182,919	53 X
FAS49777	KCP42/BC7	14	14	Y	Y (5)	5,324,139	30 X
FAS49777	KCP58/BC8	14	14	Y	Y (8)	5,214,090	64 X
FAS49777	KCX9/BC9	18	18	Y	Y (2)	4,830,983	72 X
FAS49777	KCX10/BC10	20	20	Y	Y (6)	4,939,141	66 X
FAS49777	KCX13/BC11	17	17	Y	Y (6)	5,457,230	49 X
FAS49777	KCX19/BC12	7	7	Y	Y (6)	5,023,028	29 X
FAS49777	KCX21/BC13	13	13	Y	Y (4)	5,222,183	66 X
FAS49777	KCX23/BC14	12	12	Y	Y (3)	5,166,452	63 X
FAS49777	KCX28/BC15	12	12	Y	Y (10)	5,535,701	55 X
FAS49777	KCX30/BC16	10	10	Y	Y (3)	5,071,897	53 X
FAS49777	KCX32/BC17	21	21	Y	Y (5)	5,312,130	93 X
FAS49777	KCX34/BC18	23	23	Y	Y (9)	5,428,896	75 X
FAS49777	KCX37/BC19	7	7	Y	Y (6)	5,272,156	15 X
FAS49777	KCX38/BC20	23	23	Y	Y (15)	5,761,888	88 X
FAS49777	KCX39/BC21	18	18	Y	Y (6)	5,316,083	63 X

NANOPORE run	Isolate/ Barcode	#FAST5 folders	#FASTQ folders	Miseq	Flye assembly (#contigs)	Length	COV
FAS49777	KCX40/BC22	16	16	Y	Y (2)	5,155,112	57 X
FAS49777	KCX41/BC23	16	16	Y	Y (2)	5,275,183	66 X
FAS49777	KCX48/BC24	20	20	Y	Y (2)	5,294,330	79 X
FAS49777	KCX49/BC25	8	8	Y	Y (4)	5,025,658	40 X
FAS49777	KCX60/BC26	12	12	Y	Y (2)	5,304,520	60 X
FAS49777	KCX64/BC27	10	10	Y	Y (2)	5,147,381	45 X
FAS49777	KCX69/BC28	14	14	Y	Y (7)	5,442,078	33 X
FAS49777	KCX70/BC29	25	25	Y	Y (3)	4,916,557	43 X
FAS49777	RSCP2/BC30	5	5	Y	Y (2)	4,860,202	14 X
FAS49777	RSCP6/BC31	12	12	Y	F		
FAS49777	RSCP11/BC32	16	16	Y	Y (3)	5,053,666	84 X
FAS49777	RSCP12/BC33	28	28	Y	Y (4)	5,468,099	108 X
FAS49777	RSCP17/BC34	16	16	Y	Y (12)	5,584,207	63 X
FAS49777	RSCP27/BC35	22	22	Y	Y (5)	5,370,944	56 X
FAS49777	RSCP59/BC36	13	13	Y	Y (7)	5,428,873	52 X
FAS49777	RSCP60/BC37	6	6	Y	Y (5)	5,412,491	23 X
FAS49777	RSCP62/BC38	8	8	Y	Y (4)	4,773,461	36 X
FAS49777	RSCP66/BC39	18	18	Y	Y (3)	4,877,033	73 X
FAS49777	RSCP76/BC40	7	7	Y	Y (4)	5,172,481	32 X
FAS49777	RSCX4/BC41	21	21	Y	Y (2)	4,672,015	93 X
FAS49777	RSCX8/BC42	9	9	Y	Y (2)	5,111,476	48 X
FAS49777	RSCX9/BC43	13	13	Y	Y (3)	5,107,209	38 X
FAS49777	RSCX10/BC44	20	20	Y	Y (4)	5,005,385	104 X
FAS49777	RSCX13/BC45	8	8	Y	Y (3)	4,945,615	39 X
FAS49777	RSCX18/BC46	12	12	Y	Y (3)	4,759,755	73 X
FAS49777	RSCX23/BC47	15	15	Y	Y (4)	5,004,959	65 X
FAS49777	RSCX26/BC48	9	9	Y	Y (3)	5,324,093	41 X
FAS49777	RSCX29/BC49	13	13	Y	Y (2)	4,664,392	52 X
FAS49777	RSCX31/BC50	5	5	Y	Y (5)	5,034,652	14 X
FAS49777	RSCX34/BC51	12	12	Y	Y (5)	5,040,186	45 X
FAS49777	RSCX36/BC52	22	22	Y	Y (5)	5,213,398	80 X

Nanopore flowcell ID's are given in the first column and isolate ID from Kazakhstan "starting with K" and Saudi Arabia "starting with RS". Flye assembly contig number is given in parentheses and where this number is less than 10 the genome is usually closed.

4.3.6 *fimH* analysis for *E. coli* isolates

E. coli Isolates without antibiotics from the two nations were further investigated to identify *fimH* alleles using the *fimH* typing tool at the centre for genomic epidemiology website. The most common allele found in Saudi Arabian *E. coli* isolates was *fimH34* belonging to ST46, ST58, and ST399 two each, and singletons of ST48, ST1421, ST635. Three isolates belonged to *fimH30* ST155 (2) and ST58 (1). Also, six singletons of ST46, 10,398, 744, 5359, 361 belonged to *fimH54*. The

most common alleles found in Kazakhstan *E. coli* isolates were *fimH30, 27, 54* five each. Two *fimH30* isolates belonged to ST131 and singletons of ST218, 95. Three *fimH54* isolates belonged to ST10, and singletons of ST58, 1137. Two *fimH27* isolates belonged to ST10, and singletons of ST1838, 93, 69 (**Figure 4.9**).

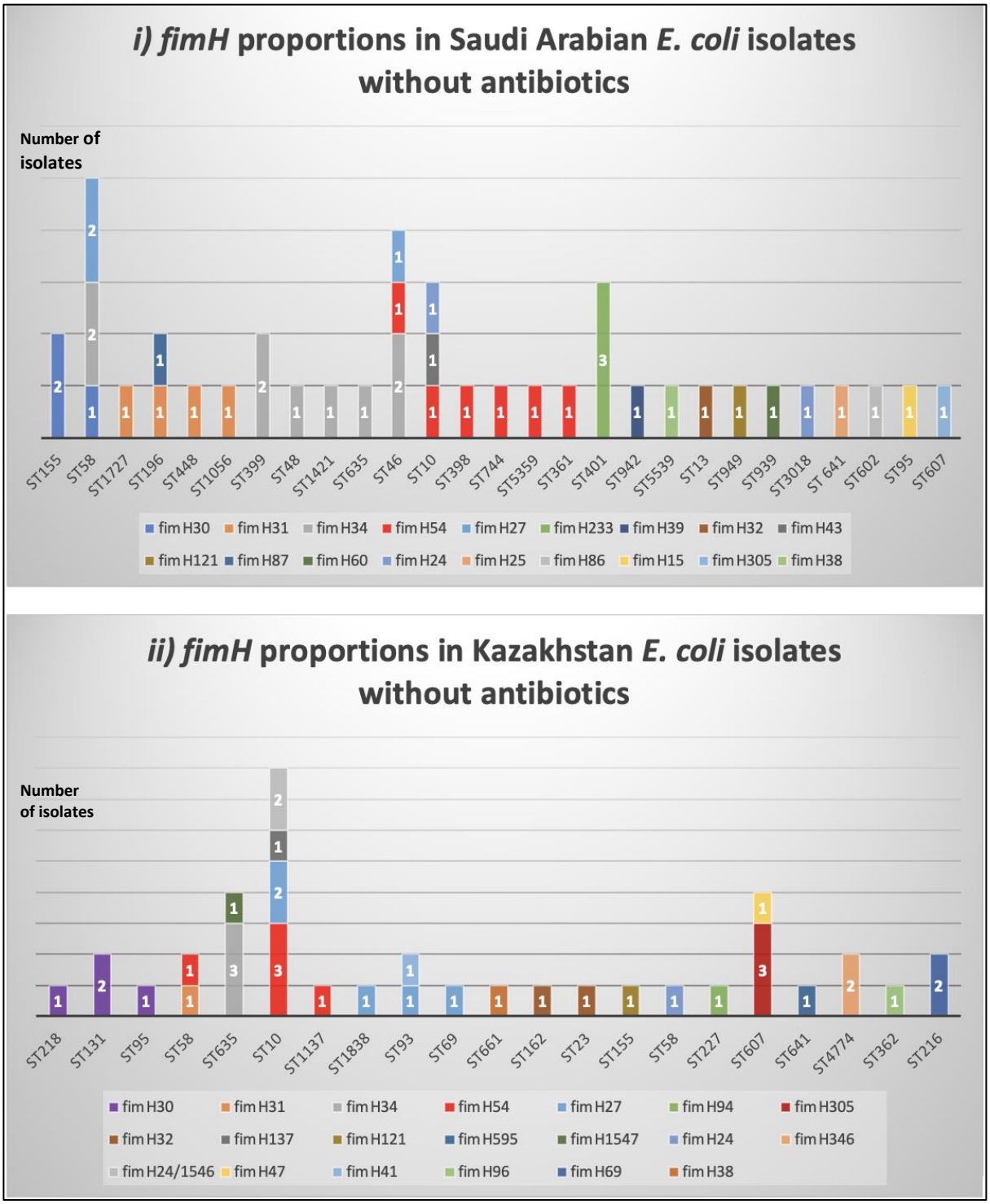


Figure 4.9. *fimH* alleles found in *E. coli* isolates without antibiotics from i) Saudi Arabia and ii) Kazakhstan MLST

The most common allele found in Saudi Arabian *E. coli* isolates selected with cefotaxime was *fimH54* belonging to ST5614 (3/7), ST10 (2/7), and singletons of ST349, ST540. Also, 3 X of *fimH24*, 3 X of *fimH28*, 3 X of *fimH215*, and a singleton of *fimH27* belonged to ST10. The most common allele found in Kazakhstan *E. coli* isolates selected with cefotaxime was *fimH30* belonging to ST131 (10/15) and ST394 (5/15). Nine isolates with *fimH54* belonged to ST10, ST162 three each and singletons of ST484, 7615, 38. Also, seven isolates with *fimH27* belonged to ST69 (4/7), ST10 (3/7) (**Figure 4.10**).

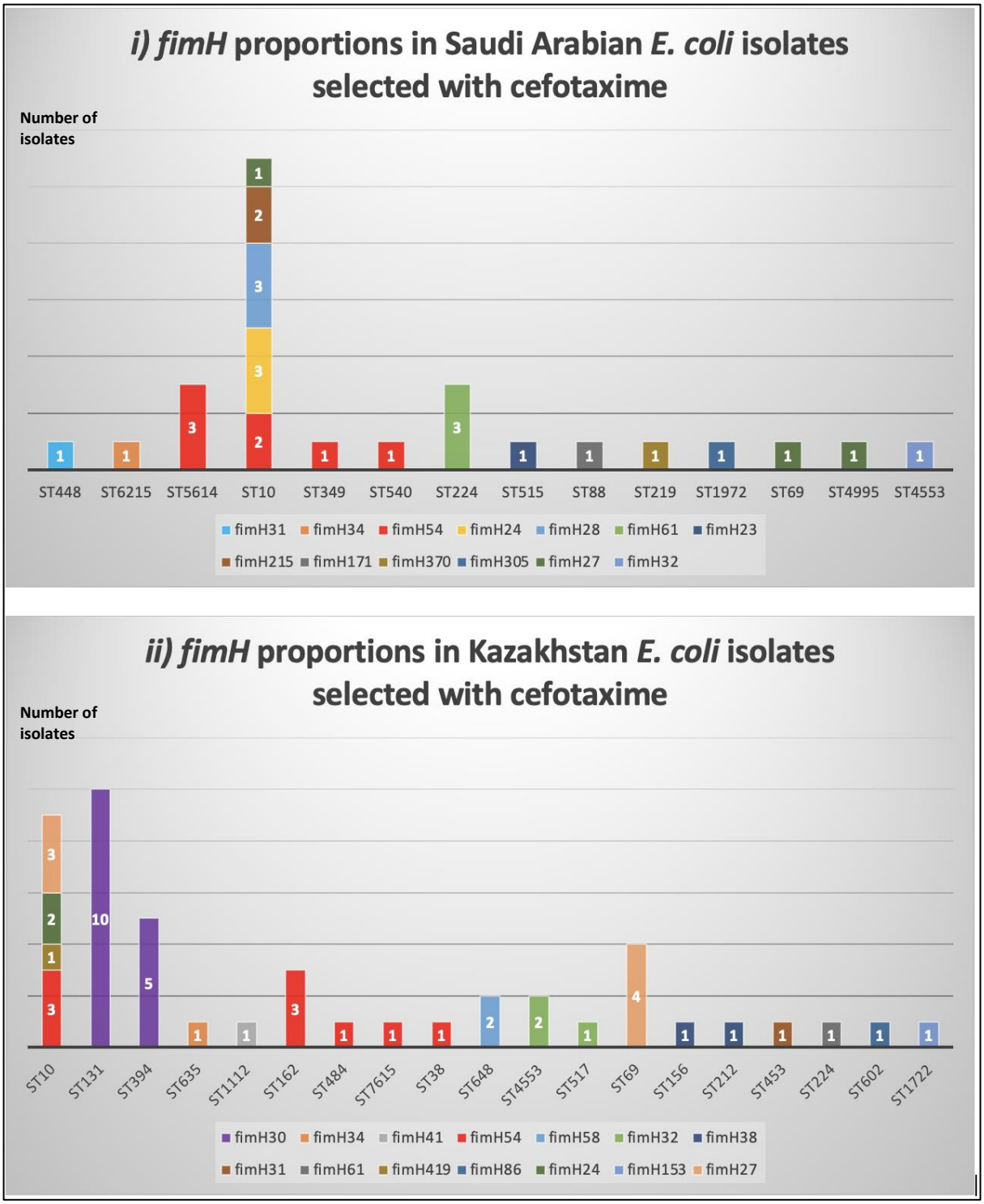


Figure 4.10. *fimH* alleles found in *E. coli* isolates selected with cefotaxime 10mg.L-1 from i) Saudi Arabia and ii) Kazakhstan MLST

Two isolates in Saudi Arabian *E. coli* isolates selected with ciprofloxacin ST10 and ST349 belonging to *fimH54* and one isolate *fimH30* belonging to ST93. The most common alleles found in Kazakhstan *E. coli* isolates selected with ciprofloxacin were *fimH30* belonging to ST131 followed by *fimH54* belonging to ST744 and ST361 three each (**Figure 4.11**).

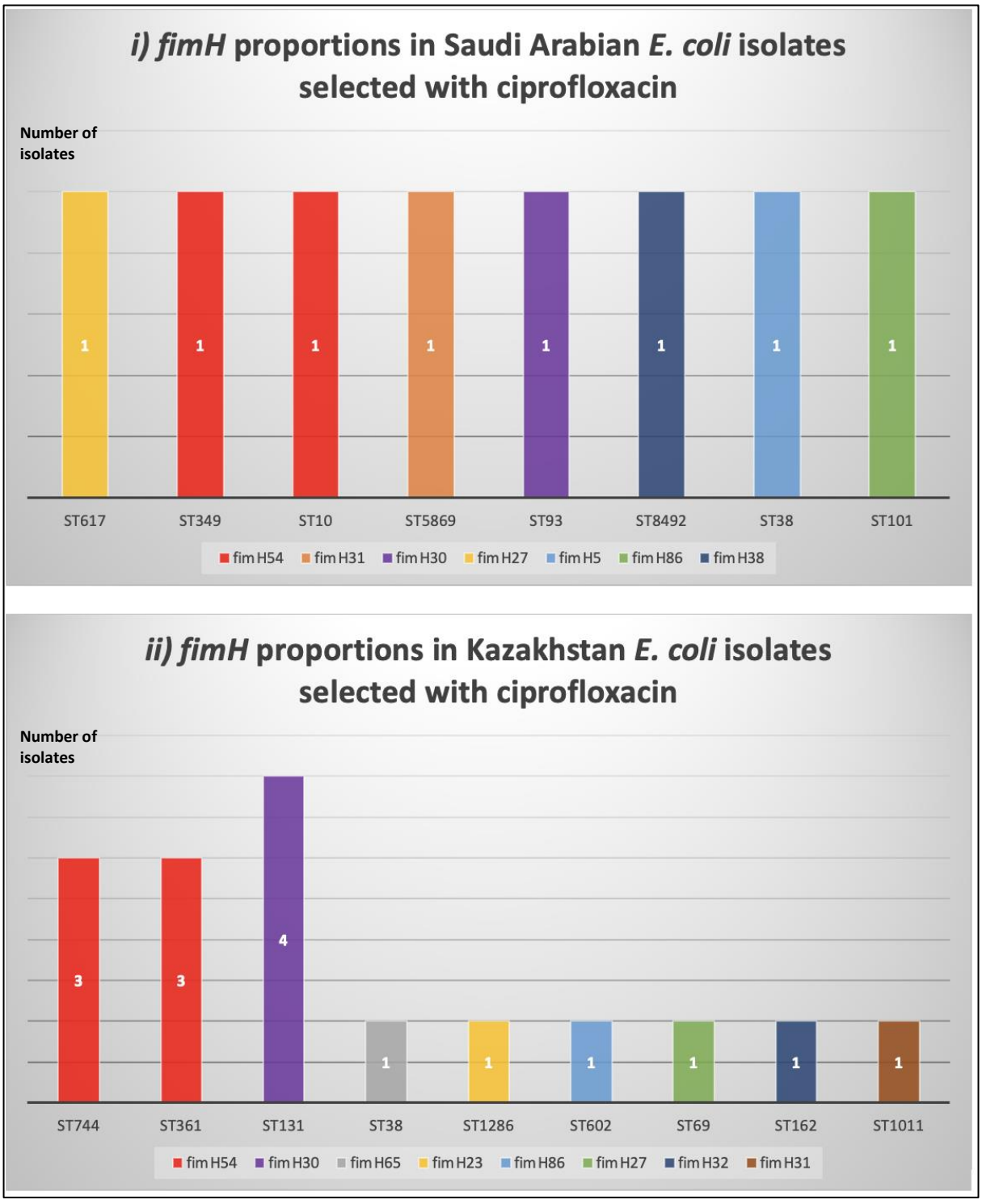


Figure 4.11. *fimH* alleles found in *E. coli* isolates selected with ciprofloxacin 1mg.L-1 from i) Saudi Arabia and ii) Kazakhstan MLST

4.3.7 Genomic location of *bla*$_{CTX-M-15}$ in antibiotic selective isolates

We located the genomic location of *bla*$_{CTX-M-15}$ by PFGE and in gel probing with *bla*$_{CTX-M-15}$ P^{32} labelled probes. The location of *bla*$_{CTX-M-15}$ gene was analyzed by PFGE with S1 nuclease (S1-PFGE) (Invitrogen Abingdon, UK). The results showed that *bla*$_{CTX-M-15}$ gene are located on a diverse range differeing from 80kb to 165kb in size (**Figure 4.12**). A few isolates (4/86) were negative for *bla*$_{CTX-M-15}$ and had lost the gene or plasmid carrying the gene during growth or storage. The isolates were divided into two groups: isolates selected with cefotaxime 10mg.L^{-1} and isolates selected with ciprofloxacin 1mg.L^{-1} for each site. For isolates selected with cefotaxime from Kazakhstan we found *bla*$_{CTX-M-15}$ on plasmids in 65% (28/43) of the isolates. These isolates were ST10 (7) and three ST394, and singletons of ST38, 162, 69, 517, 635. The *bla*$_{CTX-M-15}$ gene was found on the chromosome in 28% of isolates (12/43). Of these isolates, five were ST131. Three isolates 7%, (3/45) lost the gene or the plasmid. For isolates selected with cefotaxime from Saudi Arabia we found *bla*$_{CTX-M-15}$ on plasmids in 69% of the isolates (18/26), and 31%, (8/26) had *bla*$_{CTX-M-15}$ on the chromosome. Plasmid's locations were found in fourteen different sequence types, with 11 isolates belonging to ST10.

The isolates with ciprofloxacin from Kazakhstan found on plasmid were (78%, 7/9) belonging to ST744 (3) and two on the chromosome were belonging to ST131. Isolates with ciprofloxacin from Saudi Arabia were found on the plasmid (25%, 2/8), and five isolates were found on the chromosome (62%, 5/8). One isolate (12.5%, 1/8) lost the gene or the plasmid. One ciprofloxacin isolate from Kazakhstan that belonged to ST131 had both chromosomal and plasmid and four Saudi ciprofloxacin isolates that belonged to ST617 carried both chromosomal and two plasmids.

Plasmids ranged from 0-4 per isolate and varied in size from 50-400kb (**Figure 4.13**).

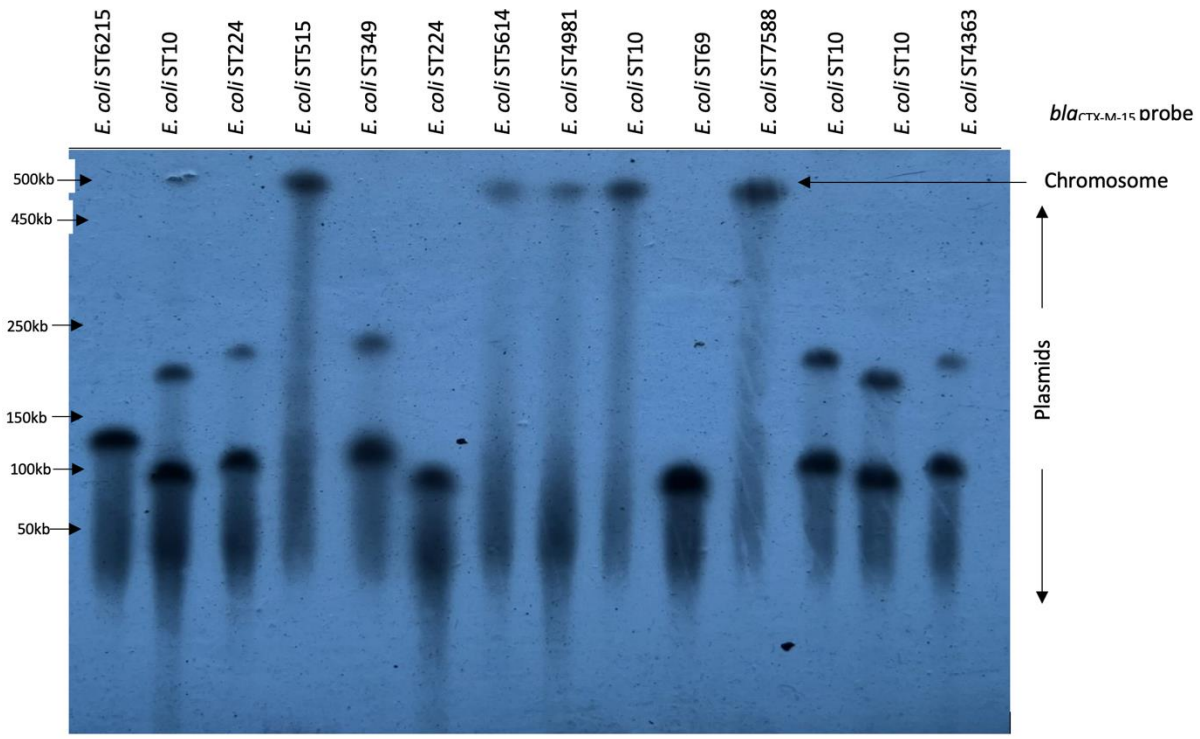


Figure 4.12. Genomic location of *bla*$_{CTX-M-15}$ gene in Saudi Arabian *E. coli* isolates

S1-Pulsed field gel electrophoresis of macro-DNA of a subset of *bla*$_{CTX-M-15}$ harbouring strains collected from Unaizah, Saudi Arabia. The *bla*$_{CTX-M-15}$ gene was carried on *incFII* plasmids in all *E. coli* strains and the sizes ranged from ~80kb to 165kb.

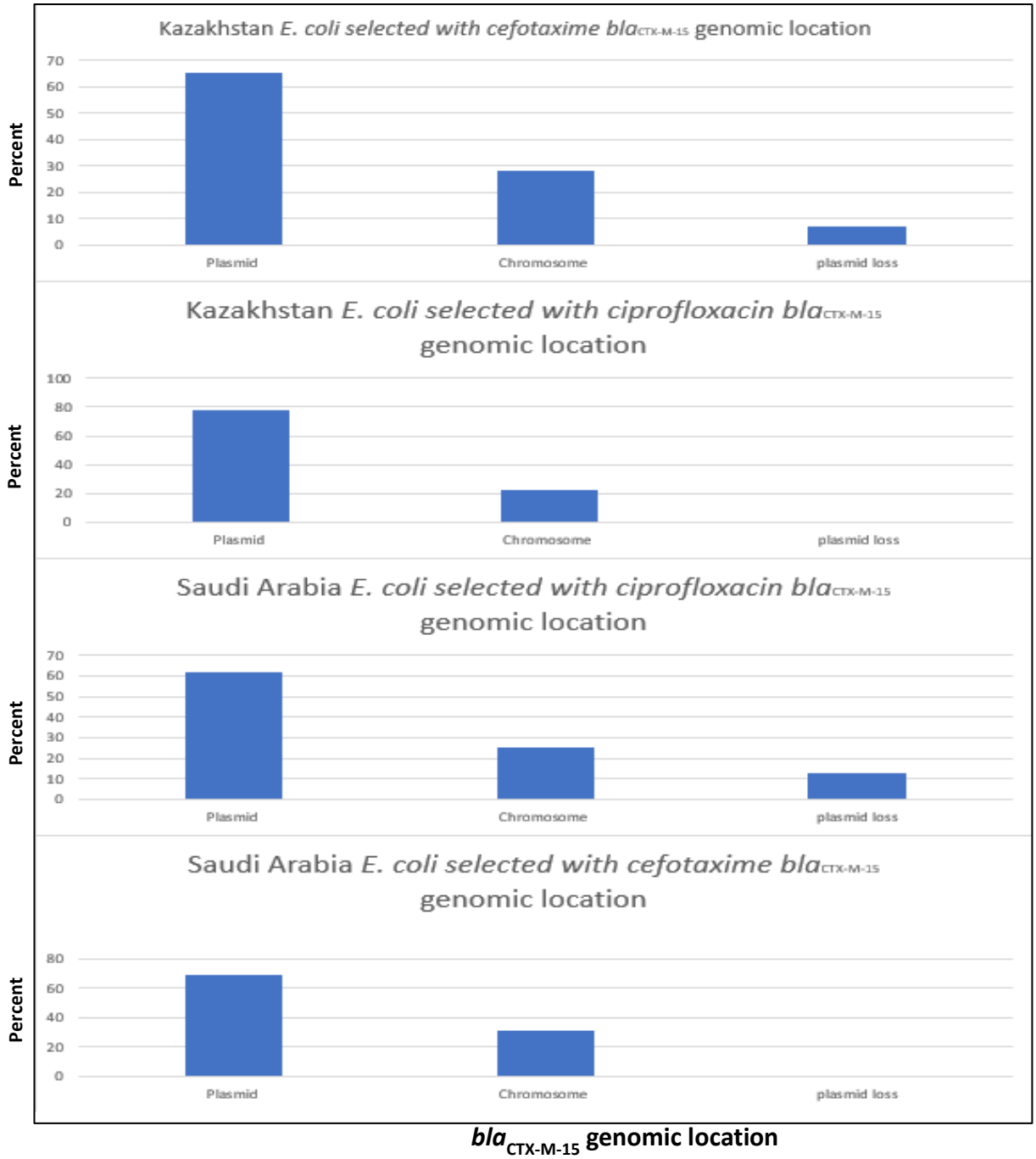


Figure 4.13. Genomic location of *bla*$_{CTX-M-15}$ in *E. coli* isolates selected with antibiotics from Karaganda, Kazakhstan and Unaizah, Saudi Arabia

4.3.8 Chromosomal insertion sites analysis

The genomic location of *bla*$_{CTX-M-15}$ genes in isolates harbouring *bla*$_{CTX-M-15}$ genes were located by local blast searches of the flye assembled contigs using Geneious software. In many cases these genomes had been closed and therefore it was easy to determine chromosomal location since the insertions were on DNA elements that were of several megabases in size. We chose 20 isolates from the two countries. This includes 5 and 8 isolates selected with cefotaxime from Saudi Arabia and Kazakhstan, respectively. Also, 1 and 3 isolates selected with ciprofloxacin were from Saudi Arabia and Kazakhstan, respectively (**Table 4.16**). Chromosomally located *bla*$_{CTX-M-15}$ genes were found in 12 different sequence types. These include 2x ST10 and a singleton of ST4995 and ST443 and one isolate could not be defined from Saudi cefotaxime isolates. Kazakhstan cefotaxime isolates were 3x ST131, 2x ST4553, 2x ST636, a singleton of ST69, ST1722, and one isolate where the MLST could not be identified. Saudi ciprofloxacin isolates were 2x ST617 and a singleton of ST38, ST648. The Kazakhstan ciprofloxacin isolates were ST131.

4.3.8.1 IS*EcP1* mediated chromosomal insertions

IS*EcP1* was responsible for (15/20), and these insertions ranged from small insertion sites of 2,911bp to very large insertion sites of >30kp. Of these isolates, seven isolates mobilised just the IS*EcP1* IS, the *bla*$_{CTX-M-15}$ resistance gene, and a small section of genes. *IS26* elements were present in eight isolates along with IS*EcP1*. *IS26* was often found flanking *bla*$_{OXA-1}$, *aacA4*, *cat*, and *tmrB* genes, suggesting that these elements were likely involved in insertion of these particular genes. Direct repeats were often found at the ends of the IS*EcP-1* transposed sections of DNA (**Figures 4.17-4.20)**.

4.3.8.2 IS*26* mediated chromosomal insertions

Of the total isolates, four isolates were mediated by the *IS26* element, a section of DNA typically mobilised by two *IS26* elements functioning as a composite transposon. IS*26* was responsible for (4/20) chromosomal insertion events. The size of the *IS26* mobilising events ranged from 5,923 bp to 12,957 bp. These elements are often associated with several resistance genes, including *bla*$_{OXA-1}$, *aacA4*, *tmrB,* and *cat* genes. Compared to the UK isolates in Chapter 3, in Kazakhstan there was no insertion found in the *molR* gene, which was a hotspot for our DH collection of isolates (**Figures 4.17-4.20**).

Table 4.13. IS*EcP1* and IS*26* mediated chromosomal insertions for Unaizah, Saudi Arabia and Karaganda, Kazakhstan *E. coli* isolates

ISOLATE ID	*E. coli* ST	Location	Figure	CTX M15	OXA -1	TEM	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT	Insertion Site detail (Where multiple insertions are found in a single strain, the insertion number is given in bold)
KCX10	NS	Karaganda		+			htrE/DR/ISEcP1/CTX-M-15/DR/htrE	2,911 bp	IS*Ec*P1	TGATA	IS*Ec*P-1 insertion into htrE is an outer membrane usher protein.
KCX9	ST636	Karaganda		+			htrE/DR/ISEcP1/CTXM-15/htrE	2,912 bp	IS*EcP1*	TGATA	IS*Ec*P-1 insertion into htrE is an outer membrane usher protein.
SCP17	ST648	Unaizah		+			insI/DR/ISEcP1/CTX-M-15/DR/insI	2,971 bp	IS*Ec*p1	TAAGA	IS*Ec*P-1 insertion into mediated insertion into insI (transposase of IS30).
SCX26	NS	Unaizah		+			ISL3/DR/ISEcP1/CTX-M15/IS26/DR/ISL3	2,984 bp	IS*Ec*P1/ IS26	AATAA	homologous recombination. 1. IS*Ecp-1* at one end and 1. IS26 at Other looks like mediated by homologous recombination.
SCX36	ST443	Unaizah		+			ybiP/DR/ISECP1/2xCTX-M-15/DR/ybiP	3,051 bp	IS*Ec*P1	TATGA	IS*Ec*p1 Insertion into ybiP inner membrane carrying 2x of CTX-M-15.
KCX60	ST1722	Karaganda		+			motA/DR/IS26/CTX-M-15/Tn3/DR/motA	5,923 bp	*IS26*	ATTTC	*IS26* mediated insert Including CTX-M-15, Tn3.
KCX69	ST4553	Karaganda		+			stpA/DR/ISEcP1/CTXM15/IS26/yokD/tmrB/ISKpn11/ IS26/DR/alaE	6,390 bp	IS*Ec*P1/ *IS26*	ACCGA	IS*Ec*P1 mediated insertion into stpA (DNA-binding protein) including phage related genes as well as IS26 bracketed regions and CTX-M-15/ tmrB. Direct repeat of ACCGA is formed on transpositional.
SCP11	ST617	Unaizah		+	+	+	rhsD/DR/ISEcP1/CTXM15/hypP/Tn3/IS26/cat/ OXA1/aacA4/IS26/DR/rhsD	8,437 bp	IS*Ec*P1/ *IS26*	TGATA	IS*Ec*P1 mediated insertion into rhsD gene.
SCP2	ST617	Unaizah		+	+	+	rhsD/DR/ISEcP1/CTXM15/hypP/Tn3/IS26/cat/ OXA1/aacA4/DR/rhsD	8,437 bp	IS*Ec*P1/ *IS26*	TGATA	IS*Ec*P1 mediated insertion into rhsD gene.
SCX23	ST10	Unaizah		+	+	+	hlyE/DR/ISEcP1/CTX-M15/hypP/Tn3/IS26/aacA4/OXA1 /catB/IS26/DR /umuD?	11,092 bp	IS*Ec*P1/ *IS26*	TATTA	IS*Ec*P1 mediated insertion into hlyE (hemolysin gene).
SCX10	ST10	Unaizah		+	+	+	hlyE/DR/ISEcP1/CTXM15/hypP/Tn3/IS26/aacA4/OXA1/ catB/IS26/DR/umuD?	11,092 bp	IS*Ec*P1/ *IS26*	TATTA	IS*Ec*P1 mediated insertion into hlyE (hemolysin gene).
KCX48	ST131	Karaganda		+			tdcA/DR/ISEcP1/CTX-M-15/hypP/Tn3/insD/insC/DR/ tdcR	11,348 bp	IS*Ec*P1	TAATA	IS*EcP-1* mediated insertion into DNA-binding transcriptional activator. Including phage related genes and CTX-M-15/Tn3/Tn2.
KCX64	ST131	Karaganda		+			tdcA/DR/ISEcP1/CTX-M15/hypP/Tn3/insD/insC/DR/ tdcR	11,384 bp	IS*Ec*P1	TAATA	IS*EcP-1* mediated insertion into DNA-binding transcriptional activator. Including phage related genes and CTX-M-15/Tn3/Tn2.
KCP21	ST131	Karaganda		+	+		hypP/IS26/CTXM15/hypP/Tn3/IS26/Tn3/IS26/tmrB/yok D/IS26/catC/OXA1/aacA4/IS26/hypP	12,902 bp	*IS26*	ND	*IS26* mediated insertion into hypothetical protein.
KCX41	ST69	Karaganda		+	+	+	WGR/DUF4132/IS26/CTXM15/hypP/Tn3/IS26/Tn3/IS26 /tmrB/yokD/IS26/catC/OXA1/aacA4/IS26/ WGR/DUF4132	12,955bp	*IS26*	ND	*IS26* mediated insert including phage related genes as well as IS26 bracketed regions Including CTX-M-15/OXA-1/catB3/aacA4/Tn3/tmrB.
KCX19	ST636	Karaganda		+	+		WGR/DUF4132/IS26/CTXM15/hypP/Tn3/IS26/Tn3/IS26 /tmrB/yokD/IS26/catC/OXA1/aacA4/IS26/WGR/DUF41 32	12,957 bp	*IS26*	ND	*IS26* mediated insert including phage related genes as well as IS26 bracketed regions Including CTX-M-15/OXA-1/catB3/aacA4/Tn3/tmrB.
KCX13	ST131	Karaganda		+	+		WGR/DUF4132/IS26/CTXM15/hypP/Tn3/IS26/Tn3/IS26 /tmrB/yokD/IS26/catC/OXA1/aacA4/IS26/WGR/DUF41 32	12,957 bp	*IS26*	ND	*IS26* mediated insert including phage related genes as well as *IS26* bracketed regions Including CTX-M-15/OXA-1/catB3/aacA4/Tn3/tmrB.
SCP12	ST38	Unaizah		+	+	+	ygfS/DR/ISEcp1/CTXM15/hypP/Tn3/IS26/cat/OXA1/aac A4/IS26/Tn3/tetR/tetA/Tn3/pemI/DR/ygfQ	18,538 bp	IS*Ecp1*/ *IS26*	TGTCA	IS*EcP-1* mediated insertion into ygfS gene (putative oxidoreductase). Insertion includes several phage related genes. it includes couple of IS26.
KCX49	ST4553	Karaganda		+		+	yjdP/DR/ISEcP1/CTX-M-15/Tn3/insD-1/pinE3/neo/ IS5075/Tn3/IS26/pine/DR/rpiB	23,490 bp	IS*EcP1*/ *IS26*	CATTA	IS*EcP-1* mediated insert including phage related genes as well as IS26 bracketed regions.
SCX34	ST4995	Unaizah		+		+	hns1/ISEcP-1/CTXM15/hypP/parM/yhdJ/yfjX/Noc2/ fimc/traM/traA/tdk	30,608 bp	IS*EcP1*	TCATA	IS*EcP-1* mediated insert into hns1 (DNA-binding transcriptional dual regulator. Very large insertion including several plasmids and phage related genes

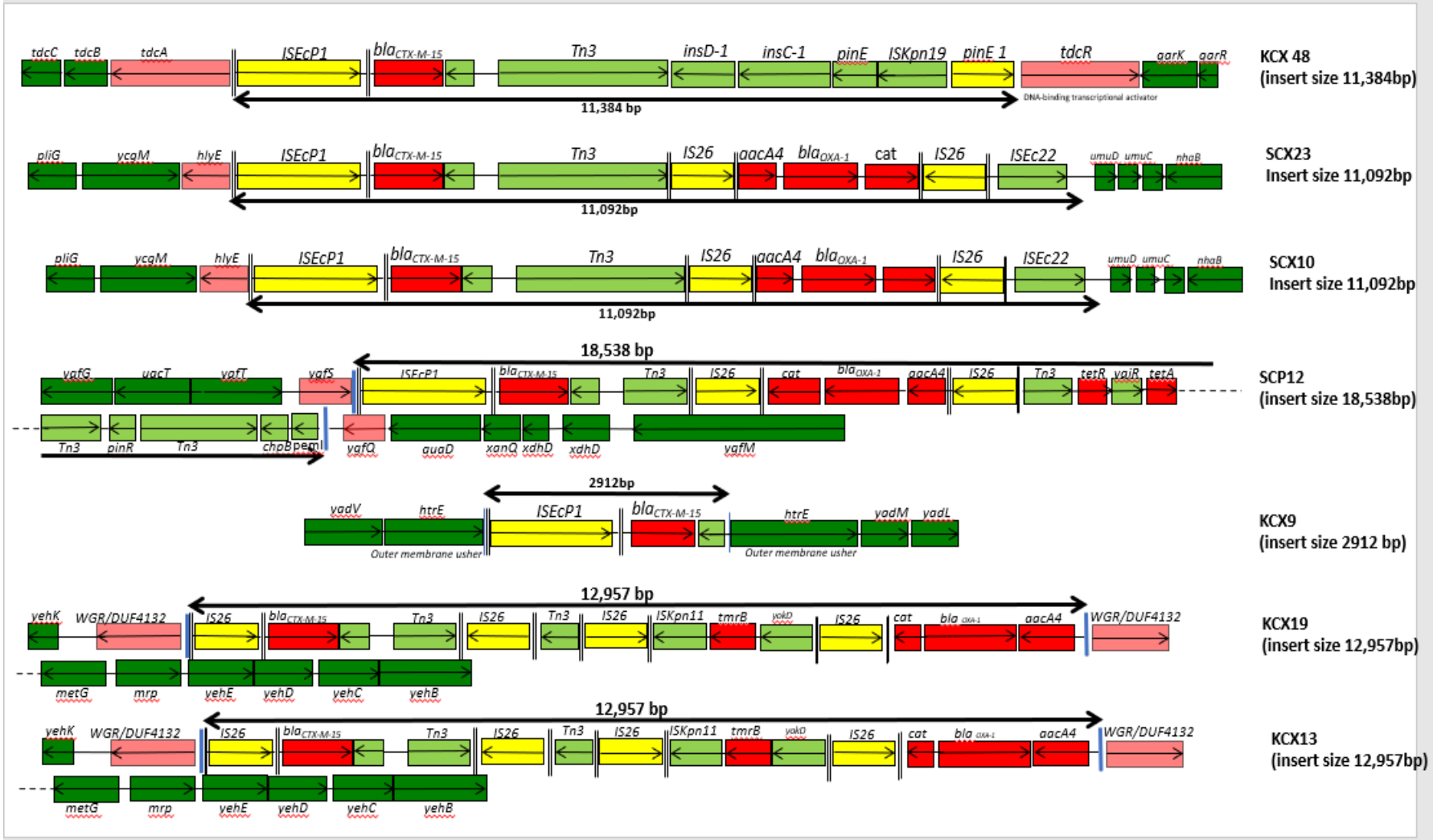


Figure 4.14. IS*EcP1* and *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

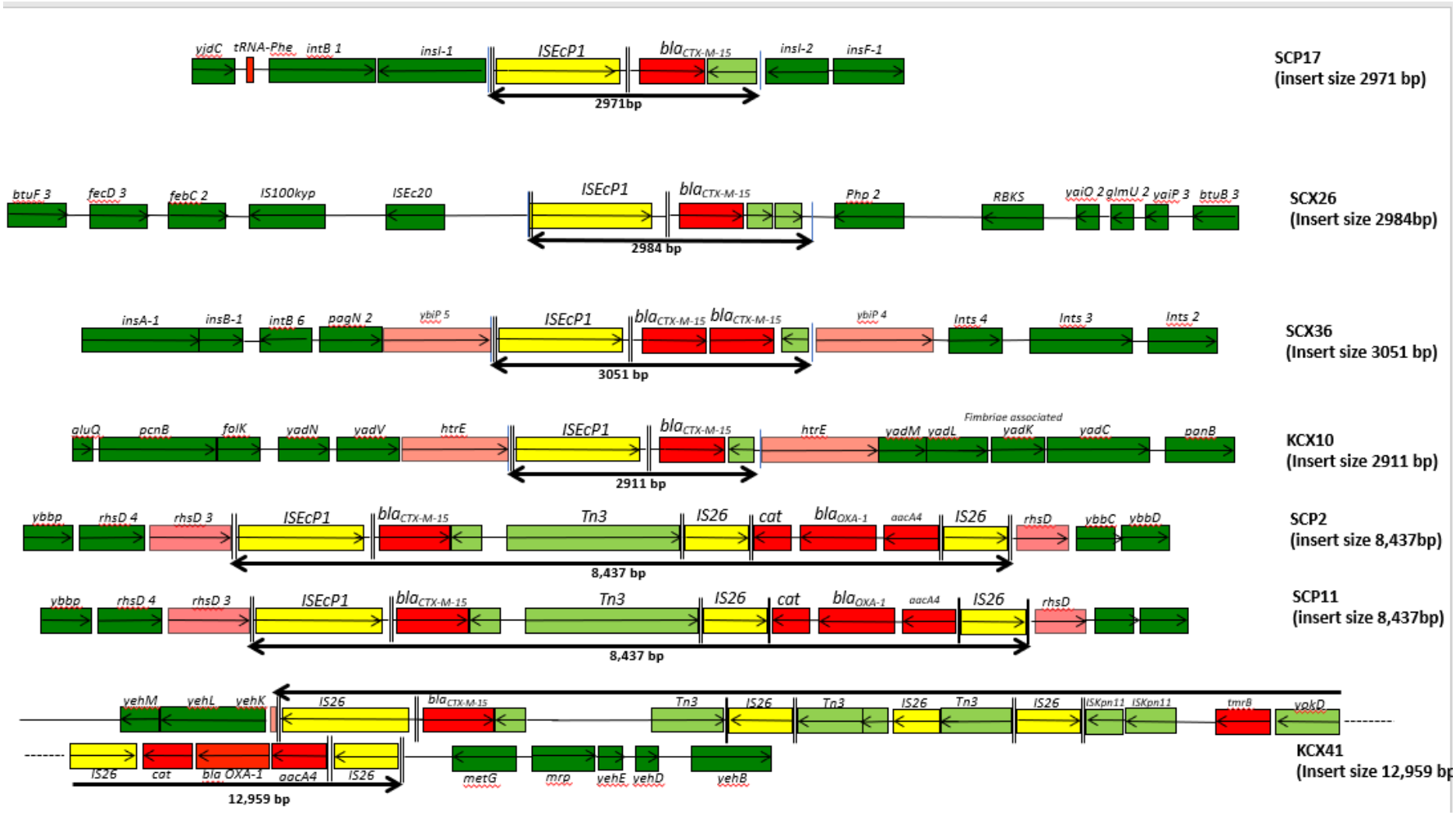


Figure 4.15 IS*EcP1* and *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

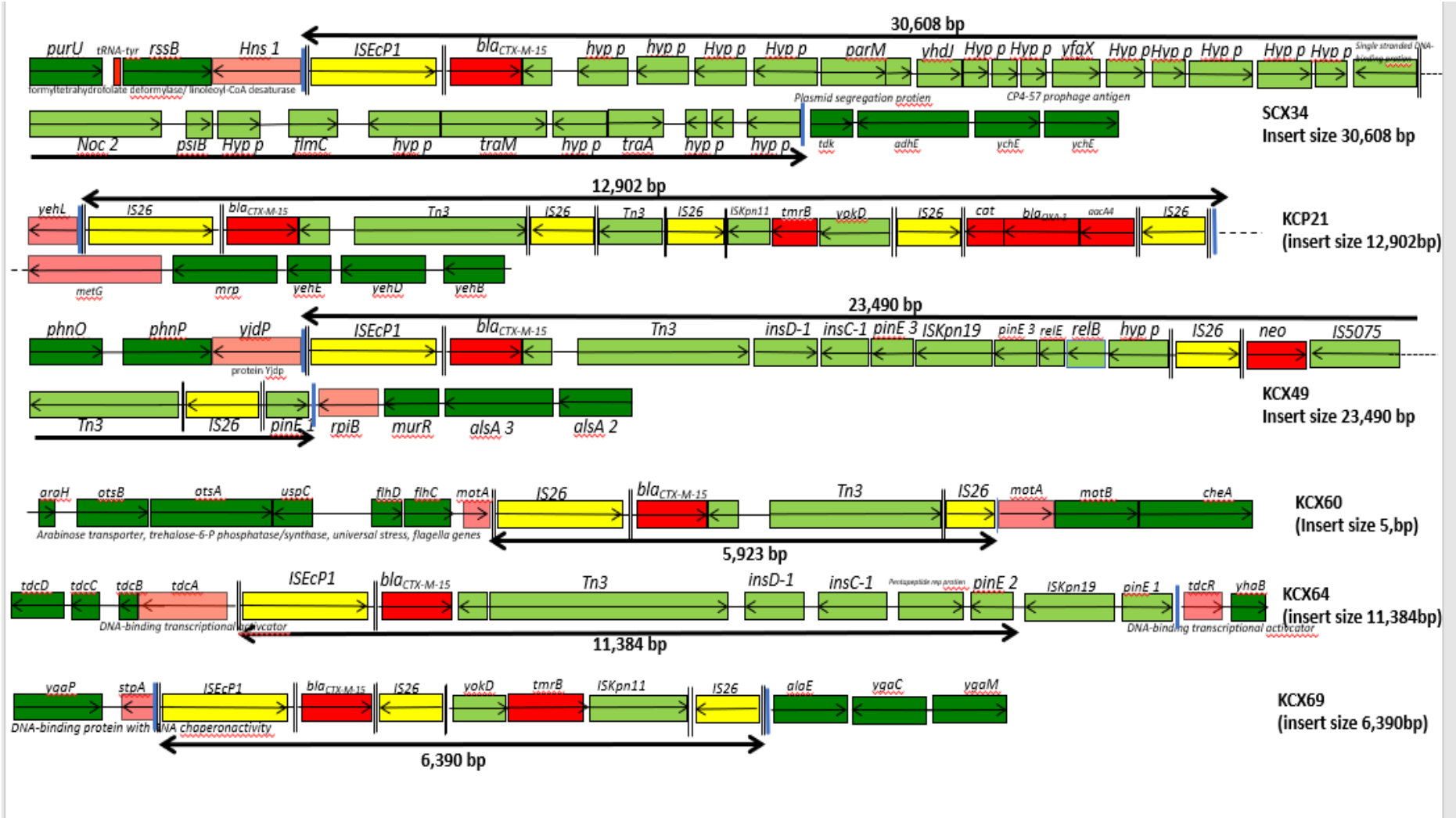


Figure 4.16. IS*EcP1* and *IS26* mediated chromosomal insertions

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

4.3.9 Plasmids analysis in Saudi Arabia and Kazakhstan isolates

The genomic location of *bla*$_{CTX-M-15}$ genes in isolates harbouring *bla*$_{CTX-M-15}$ genes were located by local blast searches of the flye assembled contigs using Geneious software. We chose 8 isolates from Unaizah that were selected with ciprofloxacin 1mg.L^{-1} and determined the chromosome and plasmid sizes and their incompatibility groups. Also, we determine the resistance genes found in the chromosome and/or in the plasmid. The *incF* incompatibility group found in five isolates (5/8) and the plasmid size ranged from 80 kb to 139 kb. The resistance genes found in both plasmids and chromosomes were presented in (**Table 4.14**).

We chose 9 isolates from Unaizah that were selected with cefotaxime 10mg.L^{-1} and determined the chromosome and plasmid sizes and their incompatibility groups. Also, we determine the resistance genes found in the chromosome and/or in the plasmid. The *incF* incompatibility group found in six isolates (6/9) and the plasmid size ranged from 80 kb to 140 kb. The resistance genes found in both plasmids and chromosome were presented in (**Table 4.14**).

We chose 5 isolates from Karaganda that selected with ciprofloxacin 1mg.L^{-1} and determined the chromosome and plasmid sizes and their incompatibility groups. Also, we determine the resistance genes found in the chromosome and/or in the plasmid. The *incF* incompatibility group found in all isolates (5/5) and the plasmid size ranged from 18 kb to 154 kb. The resistance genes found in both plasmids and chromosomes were presented in (**Table 4.15**).

We chose 12 isolates from Karaganda that were selected with cefotaxime 10mg.L^{-1} and determined the chromosome and plasmid sizes and their incompatibility groups. Also, we determine the resistance genes found in the chromosome and/or in the

plasmid. The *incF* incompatibility group was found in eleven isolates (11/12) and the plasmid size ranged from 70 kb to 168 kb. The resistance genes found in both plasmids and chromosomes were presented in (**Table 4.15**).

Table 4.14. Nanopore sequence analysis data for *E. coli* isolates from Unaizah, Saudi Arabia

Nanopore run	Isolate ID	FAST 5 folder	Flye assembly(contigs)	Coverage	Genetic location/ Incompatibility group	Chromosome and Plasmid size (kb)	Sequence type	Resistance genes
11322/FAS49777	30/RSCP2	5	Y (2)	14X	Chromo Plasmid *IncFII/IncQ1*	4,779,784 80,418	ST10	*CTX15/OXA-1/ tetA; tetC; tetR; tetD; aacA4; catB3* *TEM/ tetA; tetR*
11322/FAS49777	32/RSCP11	16	Y (3)	84X	Chromo Plasmid *IncI1-I*	4,780,322 116,015	ST69	*CTX15/OXA-1/ aacA4; cat;* *TEM/cia*
11322/FAS49777	33/RSCP12	28	Y (4)	108X	Chromo Plasmid *IncFIB/IncFII*	4,343,560 139,718	ST38	*CTX15/OXA-1/ tetA; tetR; aacA4; catB3* *TEM/ ant1; neo; tetA; tetR*
11322/FAS49777	36/RSCP59	13	Y (7)	52X	Chromo Plasmid *IncFII/IncFIA*	5,202,084 132,589	-	- *CTX15/OXA-1/ aacA4; cat; tetR; tetA; ant1;*
11322/FAS49777	37/RSCP60	6	Y (5)	23X	Chromo Plasmid *IncFIA/IncFIB/ IncQ1*	5,275,483 101,716	ST4553	- *TEM/ant1; cat; neo; tetA; tetC; tetR*
11322/FAS49777	38/RSCP62	8	Y (4)	36X	Chromo Plasmid *incY*	4,622,759 100,458	ST131	- *TEM/TEM-12/ant1; mdtL; neo; tetA; tetR*
11322/FAS49777	39/RSCP66	18	Y (3)	73X	Chromo Plasmid *IncFIB/IncFII* *incX1*	4,679,604 138,509 58,920	ST617	*TEM/ neo; ant1; tetA; tetC; tetD;* *TEM/TEM-12/ant1; tetA; tetR* *TEM/ant1*
11322/FAS49777	40/RSCP76	7	Y (4)	32X	Chromo Plasmid *IncB/O/K/Z*	4,906,936 102,037	ST617	- *CTX15/TEM/ant;*
11322/FAS49777	41/RSCX4	21	Y (2)	93X	Chromo Plasmid *IncFIB/IncFIB(K)*	4,576,857 95,158	ST617	- *CTX15/TEM/ tetA; tetR; neo;*
11322/FAS49777	42/RSCX8	9	Y (2)	48X	Chromo Plasmid *IncB/O/K/Z*	5,009,378 102,098	ST5614	- *CTX15/*
11322/FAS49777	44/RSCX10	20	Y (4)	104X	Chromo 2 plasmids *IncI1-I* *IncFIB/IncFIC*	4,689,398 114,688 140,382	ST10	- *TEM/ant;* *TEM/ ant; neo1; neo2; tetA; tetR;*
11322/FAS49777	45/RSCX13	8	Y (3)	39X	Chromo Plasmid *IncFIB/IncFIB(K)*	4,615,366 87,011	ST7588	- *CTX15/TEM/neo;*
11322/FAS49777	46/RSCX18	12	Y (3)	73X	Chromo Plasmid *IncFIB/IncFIB(K)*	4,647,740 91,619	ST7588	- *CTX15/tetA; neo;*
11322/FAS49777	47/RSCX23	15	Y (4)	65X	Chromo 2 plasmids *IncFIB/IncFIC* *IncI1-I*	4,689,335 140,023 114,667	ST10	*CTX15/OXA-1/aacA4; tetA; tetC, tetR; tetD; cat;* *TEM/ tetA; tetR; neo; ant1; mdtL; TEM/ant1;*
11322/FAS49777	49/RSCX29	13	Y (2)	52X	Chromo Plasmid *IncFIB/IncFIB(K)*	4,555,498 108,894	ST11226	- *CTX15/TEM/TEM-12/ neo; tetA;*
11322/FAS49777	50/RSCX31	5	Y (5)	14X	2 Chromo Plasmid P0111	3,767,077 1,142,850 101,984	-	*TEM/neo; tetA; tetR;* - *CTX15*
11322/FAS49777	51/RSCX34	12	Y (5)	45X	Chromo Plasmid *IncI1-I*	4,842,666 112,509	ST5614	*CTX15/* *TEM/ ant1;*

RS In the isolate ID stands for Saudi Arabian *E. coli* isolates and CP stands for isolates selected with ciprofloxacin 1 mg.L^{-1} and CX stands for isolates selected with cefotaxime 10 mg.L^{-1}. Each FAST5 folder contains so many reads which is generally 4000 reads. Y stands for yes; the genome was closed, and the number represents the closed DNA elements and where this number is less than 10 the genome is usually closed. Mostly a single chromosome and number of plasmids and in the last column we presented all the resistance genes were found in these locations and (-) means no resistance genes were found.

Table 4.15. Nanopore sequence analysis data for *E. coli* isolates from Karaganda, Kazakhstan

Nanopore run	Isolate ID	FAST5 folder	Flye assembly (contigs)	Coverage	Genetic location/ Incompatibility group	Chromosome and Plasmid size (kb)	Sequence type	Resistance genes
11322/FAS49777	1/KCP9	13	Y (2)	57X	Chromo Plasmid *IncFIB*	4,846,295 120,671	ST1286	- *TEM/ tetA; tetR;*
11322/FAS49777	3/KCP18	21	Y (2)	126X	Chromo Plasmid *IncFII*	4,788,648 89,350	ST744	*TEM/neo; ant1; tetA; tetC; tetD* *CTX15/ant1;*
11322/FAS49777	5/KCP35	10	Y (5)	38X	Chromo Plasmid *IncFIA/IncFIB/* *IncFII*	5,024,246 154,741 115,391	ST131	- *TEM/ ant1; neo; tetA; tetR;* *CTX15/dfrD;*
11322/FAS49777	7/KCP42	14	Y (5)	30X	Chromo Plasmid *IncFIA/IncFIB/* *IncFII*	5,024,146 154,745 115,386	-	- *TEM/ ant1; neo; tetA; tetR* *CTX15/dfrD*
11322/FAS49777	8/KCP58	14	Y (5)	64X	Chromo Plasmid *IncFIA/IncFIB/* *IncFII*	4,954,463 142,536 18,986	ST131	- *TEM/neo; tetA; tetR* *CTX15/*
11322/FAS49777	14/KCX23	12	Y (3)	63X	Chromo Plasmid *IncFII*	5,025,426 70,950	ST131	*TEM/TEM-12/tetA; tetR; cat; neo* *CTX15/*
11322/FAS49777	15/KCX28	12	Y (10)	55X	Chromo Plasmid *IncFII*	3,240,956 93,337	ST10	- *CTX15/ant1;*
11322/FAS49777	16/KCX30	10	Y (3)	53X	Chromo 2 Plasmid *p0111* *IncFIB(K)/IncN*	4,872,025 76,377 123,495	ST10	- *CTX15/ant1; tetA; tetR* *CTX15/TEM/neo; tetA; tetR*
11322/FAS49777	17/KCX32	21	Y (5)	93X	Chromo 3 Plasmid *incX* *IncFIB/IncFIB(K)* *IncX*	5,054,048 54,852 124,645 49,901	ST131	- *TEM/ant1;* *CTX15/TEM/tetA; tetR; neo* *TEM/ ant1; mdtL;*
11322/FAS49777	18/KCX34	23	Y (9)	75X	Chromo Plasmid *Col156/IncFIA/IncFIB/IncFII*	4,405,535 173,665	ST636	- *CTX15/TEM/TEM-12/ ant1; tetA; tetR; tmrB; neo*
11322/FAS49777	19/KCX37	7	Y (6)	15X	Chromo Plasmid *IncFIB*	5,055,003 124,646	ST69	- *CTX15/TEM/TEM-12/ neo; tetA; tetR*
11322/FAS49777	20/KCX38	23	Y (15)	88X	Chromo Plasmid *IncFII*	3,280,672 76,826	ST394	- *CTX15/*
11322/FAS49777	21/KCX39	18	Y (6)	63X	Chromo 3 Plasmids *IncFIB/IncFIB(K)* *incX1* *incX1*	5,054,068 124,636 49,880 54,855	ST636	- *CTX15/TEM/tetA/tetR/* *TEM/ ant1* *TEM/ant1;*
11322/FAS49777	22/KCX40	16	Y (2)	57X	*Chromo* *Plasmid* *IncI1-I (Alpha)*	5,065,060 90,052		- CTX15/
11322/FAS49777	24/KCX48	20	Y (2)	79X	Chromo Plasmid *IncFIA/IncFIB/IncFIC*	5,128,788 165,542	ST10	CTX15/ *TEM/ neo;*
11322/FAS49777	28/KCX69	14	Y (7)	33X	Chromo Plasmid *IncFIA/IncFII*	5,274,869 103,255	ST69	*CTX15/OXA-1/tetA; tetR; tmrB; cat; aacA4;* *TEM/*
11322/FAS49777	29/KCX70	25	Y (3)	43X	Chromo Plasmid *IncFIA/IncFIB/IncFII*	4,740,911 168,031	ST394	- *CTX15/OXA-1/cat; aacA4; tetA; tetR; tetC; tmrB;*

K In the isolate ID stands for Kazakhstan *E. coli* isolates and CP stands for isolates selected with ciprofloxacin 1 mg.L^{-1} and CX stands for isolates selected with cefotaxime 10 mg.L^{-1}. Each FAST5 folder contains so many reads which is generally 4000 reads.

Table 4.16. IS*26* and IS*EcP1* mediated plasmid insertions in Saudi Arabia and Kazakhstan *E. coli* isolates

ISOLATE ID	*E. coli* ST	Location	CTX M15	OXA -1	TEM	Adjacent genes/Inserted Genes/mobile element DR = direct repeat flanking insert	Insert size	DRIVER	DIRECT REPEAT
SCP59	NS	Unaizah	+	+		nqrC/IS*26*/hypP/tap/hypP/IS*26*/sprC/folP/emrE/ant1/folA/xerC/hypP/IS*26*/bla_{CTXM15}/hypP/Tn3/IS*26*/Tn3/IS*26*/cat/bla_{OXA1}/aacA4/IS*26*/pinR/Tn3/hypP/tetA/yajR2/tetR/Tn3/IS*26*/Tn3/chpB/pemI/hypP	32,002 bp	IS*26*	ND
SCX18	ST7588	Unaizah	+			mlrA/hypP3x/IS*26*/pinE/xerC/folA/IS*26*/Tn3/IS5075/hypP2x/neo/hypP3x/*tetR*/yajR/*tetA*/hypP2x/bla_{TEM}/$bla_{TEM\ 12}$/pinE/Tn3/DR/IS*EcP1*/insE/insEF/bla_{CTXM15}/Tn3/insD/insC/pinE/ISKpn19/pinE/relE/relB/hypP/IS*26*/Tn3/insA/DR/evgS	39,708 bp	*IS26* &IS*EcP1*	TCATA
SCX29	ST11226	Unaizah	+			mlrA/hypP3x/IS*26*/hypP2x/IS5075/hypP 2x/neo/hypP3x/tetR/yajR/tetA/hypP2x/TEM/pinE/Tn3/DR/IS*EcP1*/insE/insEF/bla_{CTXM15}/Tn3/insD/insC/pinE/ISKpn19/pinE/insA/DR/evgS	27,600 bp	IS26 &IS*Ec*p1	TCATA
SCX31	NS	Unaizah	+			hypP/IS*EcP1*/$bla_{CTX\text{-}M\text{-}15}$/hypP	2,499 bp	IS*Ec*P1	ND
SCX4	ST617	Unaizah	+		+	mlrA/hypP3x/IS*26*/hypP/pinE/xerC/folA/IS*26*/betT/insAB/insA/Tn3/pinR/ychM/IS5075/hypP2x/nemR/nemA/hypP3x/fabG/malY/IS5075/hypP/neo/hypP3x/tetR/tetA/tetA/hypP2x/*bla*TEM/pinE/Tn3/DR/IS*EcP1*/insE/insEF/bla_{CTXM15}/Tn3/insD/insC/pinE/ISKpn19/pinE/relE/relB/hypP/IS*26*/Tn3/insA/DR/evgS	53,858 bp	IS26 &IS*Ec*P1	TCATA
SCX8	ST5614	Unaizah	+			hypP/IS*EcP1*/bla_{CTXM15}/Tn3/insD/insC/insC/pinE/ISKpn19/pinE/hypP2x/IS*26*/hypP/cia/hypP	14,970 bp	*IS26 &IS*EcP1*	ND
KCP18	ST744	Karaganda	+			stpB/IS*26*/hypP/bla_{CTXM15}/Tn3/insD/insC/incC/pinE/ISKpn19/pinE/hypP2x/IS*26*/hypP/relE/relB/betT/betT/flxA/IS*26*/tap/IS600/srpC/folP/ermE/ant1/folA/xerC/IS*26*/Tn3/chpP/pemI/hypP	32,453 bp	*IS26*	ND
KCP35	ST131	Karaganda	+			tRNA/hypP/DR/IS*EcP1*/insAB/insA/bla_{CTXM15}/hypP10x/DR/hypP 7x	24,153 bp	IS*Ec*P1	TCGTA
KCP42	-	Karaganda	+			tRNA/hypP/DR/IS*EcP1*/insAB/insA/bla_{CTXM15}/hypP10x/DR/hypP 7x	24,153 bp	IS*Ec*P1	TCGTA
KCX23	ST131	Karaganda	+			stpB/IS*26*/hypP/bla_{CTXM15}/Tn3/insD/insC2x/pinE/ISKpn19/pinE/hypP2x/IS*26*/hypP/relE/relB/betT/betT/flxA/IS*26*/tap/IS600/srpC/folP/ermE/ant1/folA/xerC/IS*26*/Tn3/chpP/pemI/hypP	32,453 bp	*IS26*	ND
KCX28	ST10	Karaganda	+			stpB/IS*26*/hypP/bla_{CTXM15}/Tn3/insD/insC/incC/pinE/ISKpn19/pinE/hypP2x/IS*26*/hypP/relE/relB/betT/betT/flxA/IS*26*/tap/IS600/srpC/folP/ermE/ant1/folA/xerC/IS*26*/Tn3/chpP/pemI/hypP	32,453 bp	*IS26*	ND
KCX30	ST10	Karaganda	+		+	folA/IS*26*/Tn3/IS5075/folP/neo/bla_{TEM}/pinE/Tn3/IS*EcP1*/bla_{CTXM15}/Tn3/insD/insC/pinE/ISKpn19/pinE/relE/relB/IS*26*/hypP/Tn3/*tetR*/yaj/*tetA*/hypP/xerD/rcsC/virB4x2/ptlE/virB9/virB10/virB11/topB/hns/clsB/traC/ISKox/umuD/umuC/hypP/sopB/repB/blc/mlrA/hypP3x/IS*26*/ISVsa/yghA/pspF/emrE/folP	73,721 bp	IS*Ec*P1	TAATA
KCX34	ST636	Karaganda	+		+		52,456 bp	IS26 & IS*Ec*P1	ND
KCX38	ST394	Karaganda	+			stbB/hypP3x/IS*26*/bla_{CTXM15}/Tn3/insD/insC/pinE/ISKpn19/pinE/relE/relB/IS*26*/Tn3/IS*26*/Tn3/chpB/pemI/hypP3x/hha	23,167 bp	*IS26*	ND
KCX39	ST636	Karaganda	+		+		82,111 bp	*IS26 & ISEcP-1*	
KCX40	-	Karaganda	+		+	umuC/IS*EcP1*/Tn3/insD/insC2x/pinE/ISKpn19/IS*26*/hypP4x	12,167 bp	*IS26 & ISEcP-1*	ND
KCX70	ST394	Karaganda	+	+		hypP/IS*26*/pinQ/nemR/hypP/nemA/hypP/fabG/malY/hypP/IS4321/agb/resA/ybbA/hypP/tbd/nqrC/resA/hypP/IS*26*/insAB/ISVsa/*tetD*/*tetC*/*tetA*/*tetR*/hypP3x/insAB/IS*26*/hypP3x/tap/IS*26*/folP/IS*26*/IS*EcP1*/bla_{CTXM15}/Tn3/IS*26*/ISKpn11/*tmrB*/yokD/IS*26*/aacA4/bla_{OXA1}/cat/IS*26*/Tn3/chpB/pemI/hypP	54,340 bp	*IS26 & ISEcP-1*	ND

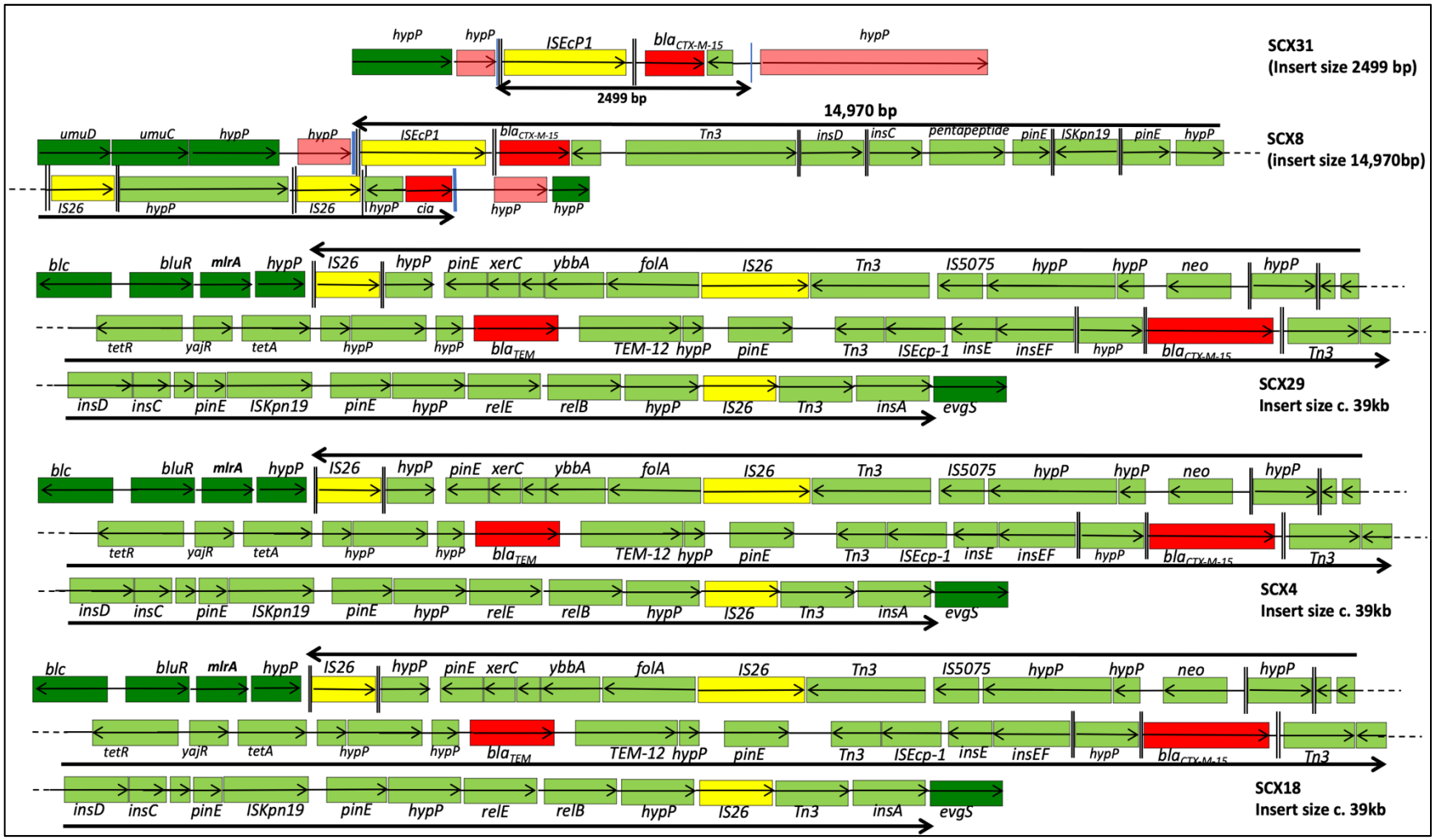


Figure 4.17. *IS26* and *ISEcP-1* mediated Plasmid insertions for Saudi Arabian *E. coli* selected for cefotaxime

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*CTX-M-15 gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

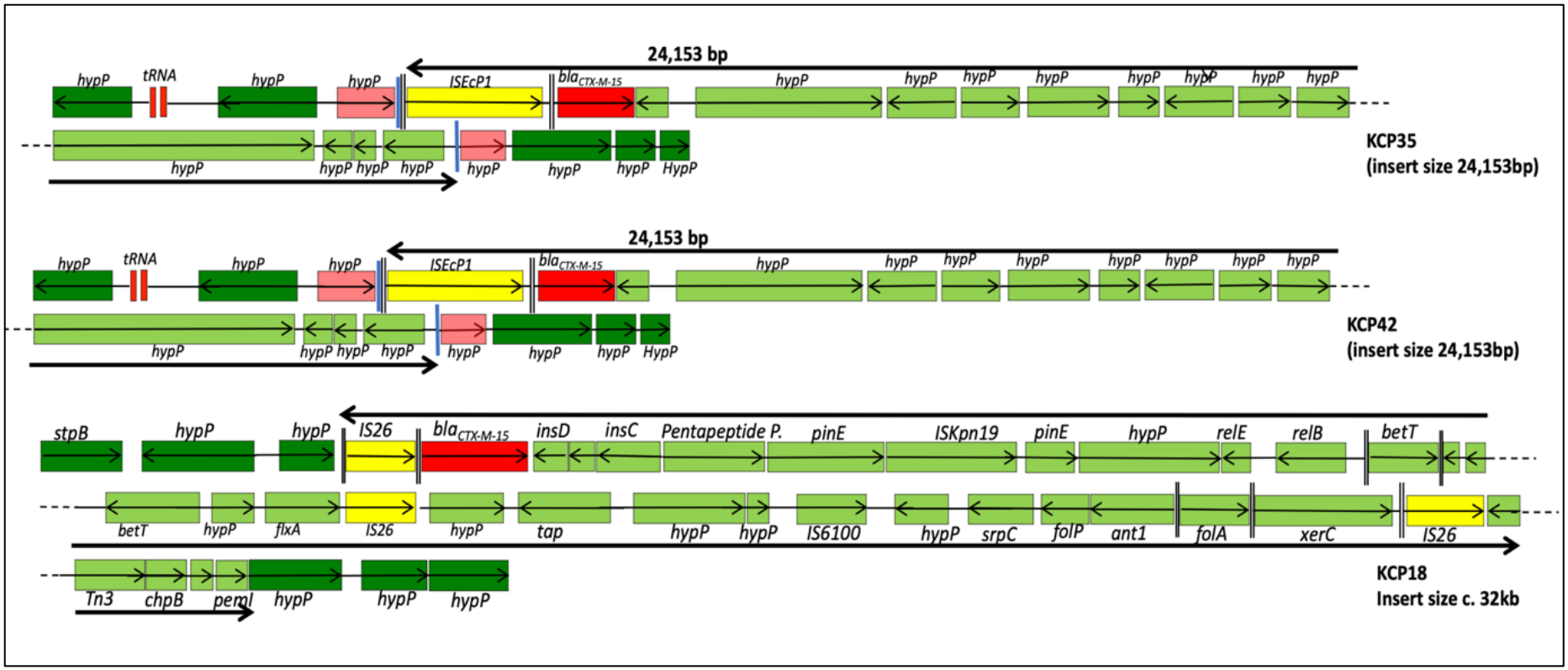


Figure 4.18. *IS26* and *ISEcP-1* mediated Plasmid insertions for Kazakhstan *E. coli* selected for ciprofloxacin

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{CTX-M-15}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

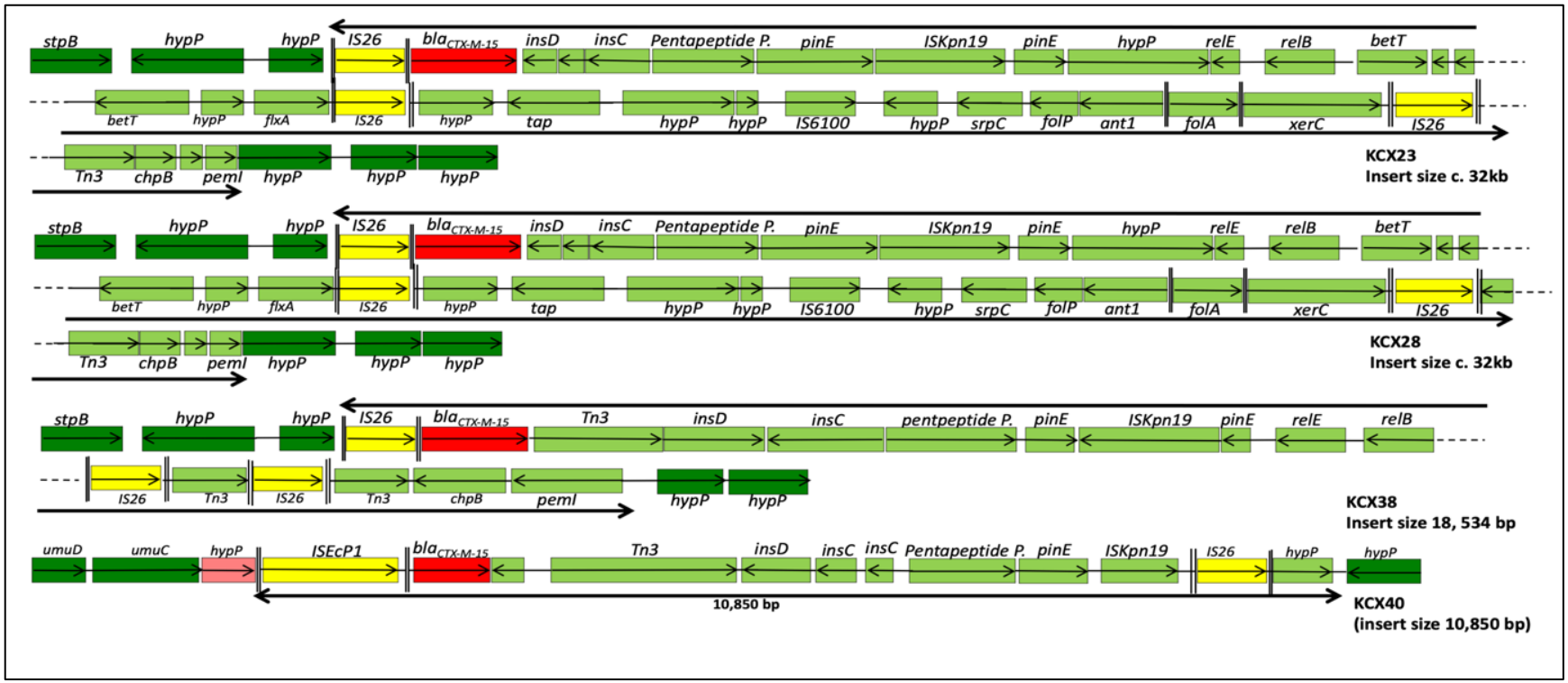


Figure 4.19. *IS26* and *ISEcP-1* mediated Plasmid insertions for Kazakhstan *E. coli* selected for cefotaxime

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively.

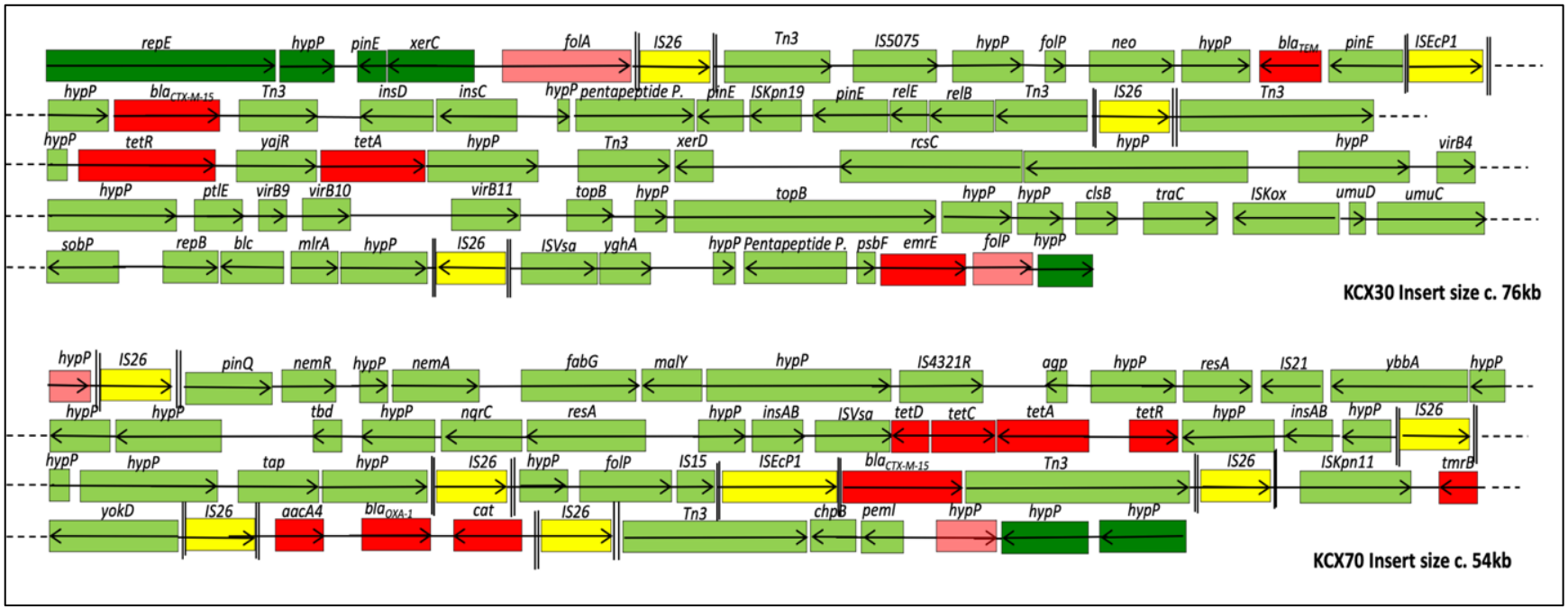


Figure 4.20. IS*26* and IS*EcP-1* mediated Plasmid insertions for Kazakhstan *E. coli* selected for cefotaxime

Genes are indicated by blocks with arrows indicating the direction of their transcription. Dark green blocks indicate chromosomal genes. Pink-coloured blocks are chromosomal genes that are disrupted by the insertion of the *bla*$_{\text{CTX-M-15}}$ gene (coloured in red). Light green blocks are co-transposed genes. Mobile elements involved in the transposition event are coloured yellow with their inverted repeats represented as double vertical lines. Blue vertical lines indicate the position of direct repeats flanking the insertion site. Isolate names in black and blue indicate bacteraemia and faecal isolates, respectively

4.4 Discussion

Phylogenetic analysis was applied to *E. coli* isolates from two different sites in Saudi Arabia and Kazakhstan by Clermont PCR, and whole genome sequence using Illumina MiSeq. In this chapter, we sought to examine the diversity of phylogroups, MLST, and the genetic basis of antibiotic resistance between different nations. MLST data of *E. coli* isolates collected from Unaizah, Saudi Arabia showed that ST58 and ST10 were the most frequently found STs among isolates without antibiotic selection. These results are so important since no other research group in Unaizah city has ever investigated the sequence types of *E. coli* in the area. Other clinical studies as shown in (**Table 4.1**) have found that ST131 is the predominant strain in Saudi Arabia. These results are different from those found in Pakistan sewage samples, where ST394 was the predominant sequence type followed by ST10 (Zahra et al. 2018), and also different from those found in Europe and North America, where ST131 was the predominant sequence type. Phylotype analysis indicates that most isolates from Unaizah sewage samples belonged to the non-pathogenic groups A and B1, and none of the isolates from Unaizah belonged to phylogroup B2. These results were similar to those of the isolates collected from Pakistan, where only one isolate belonged to Phylogroup B2 (Zahra et al. 2018). However, the most frequently found STs in Karaganda was ST10, and three isolates belonged to phylogroup B2, two of which were ST131, and one was ST95. *E. coli* isolates without antibiotics enabled us to measure the resistance mechanism in these species in both countries. We were able to measure the spread of resistance mechanisms throughout *E. coli* by isolating the bacteria without antibiotic selection. There is widespread distribution of the *bla*$_{\text{CTX-M-15}}$ gene worldwide, but it was not found in Unaizah *E. coli* isolates from sewage. However, the prevalence of *bla*$_{\text{TEM-1}}$

and *bla*$_{OXA-1}$ was documented in this area at 20% and 5%, respectively (**Table 4.4**). In a clinical study conducted in a Riyadh hospital, they found the prevalence of the *bla*$_{CTX-M-15}$ gene in *E. coli* was 60% (Al-Agamy et al. 2018). Comparing Kazakhstan, we found the *bla*$_{CTX-M-15}$ gene colonising 5% of *E. coli* isolates from sewage without antibiotics. However, the prevalence of *bla*$_{TEM-1}$ was 18%, and *bla*$_{OXA-1}$ was absent (**Table 4.9**). It is interesting that other resistance genes, i.e., *sul, dfrA, aph(3")-Ib*, and *aph(6)-Id*, were almost doubled in Unaizah compared to Karaganda, with the exception of *aadA* (**Table 4.8**).

When selecting *E. coli* with antibiotics, the phylogroups change dramatically to pathogenic types. We found the prevalence of the *bla*$_{CTX-M-15}$ gene in *E. coli* with cefotaxime 10 mg/L selection was 69% and 71% in Unaizah and Karaganda, respectively (**Tables 4.5 and 4.10**).

Our results of antibiotic discs diffusion for Unaizah, SA, and Karaganda, KA, showed that most pathogenic phylogroups B2 and D were more resistant (**Figures 4.7 and 4.8**). This means that increasing antibiotic usage is likely to drive an increase in the prevalence of the B2 phylogroup in the community. For example, when selecting an isolate with cefotaxime or ciprofloxacin, the phylotype switches immediately to pathogenic types, allowing them to take over. Also, the downside of adding antibiotics is that it allows pathogenic *E. coli* to take over the commensal ones. This data is so important since no other research group has ever shown such an important result before.

Analysis of *fimH* data revealed that several isolates were *fimH* null. However, we identified three *fimH30* alleles in *E. coli* isolates without antibiotics from Unaizah that belonged to ST155 (2) and ST58 (**Figure 4.9**), but interestingly, they were absent in *E. coli* isolates with cefotaxime selection from the same site and replaced by *fimH54* 25% (7/28) (**Figure 4.10**). However, we found that *fimH30* was present in Karaganda

E. coli isolates with cefotaxime selection of 32% (15/47), specifically with ST131 (10/15) and ST394 (5/15) (**Figure 4.10**). *fimH30* was found in most isolates that were resistant to quinolones and cephalosporins, and it was used to identify if the isolates shared common alleles known to promote strong building to epithelial cells. Analysis of short read indicates that most plasmids from Unaizah, SA, and Karaganda, KA, belong to the *incF* incompatibility group. The *incF* incompatibility group carries more resistance genes compared with other incompatibility groups. Plasmid typing confirmed that horizontal gene transfer through *IncF* plasmid exchange was likely also involved in the spread of the *bla*$_{\text{CTX-M-15}}$ strain. These plasmids can be thought of as pandemics because they have been found in bacteria from different countries and with different origins. The occurrence of these plasmid types is closely linked to the positive selection that is exercised using antibiotics. As a result, the prevalence of these plasmid types has increased in comparison to that of bacterial populations that have not been preselected for antimicrobial resistance. Moreover, their presence is connected to virulence factors, which are variables that contribute to the fitness of their bacterial host (Sherley et al. 2003; Karisik et al. 2006). We looked at where the *bla*$_{\text{CTX-M-15}}$ gene was on the chromosome using PFGE and radioactive probing to search the whole collection for chromosomal and plasmid forms of the gene. The results from sewage samples showed that the *bla*$_{\text{CTX-M-15}}$ gene on the plasmid is higher in both countries (**figure 4.13**). This is different from the data presented in the UK, which shows that chromosomal carriage of *bla*$_{\text{CTX-M-15}}$ is hugely important, specifically in bacteraemia collection.

In summary, we demonesrtated the the diversity of phylogroups, MLST, and the genetic basis of antibiotic resistance in two different nations.

Chapter 5 The Link between Sepsis Rates and Carriage of Pathogenic Sequence Types in the UK Community

5.1 Introduction

5.1.1 Prior *E. coli* phylogroup studies

To appreciate the significance of the data I will present in this chapter, allow me to first put this research into the context of prior *E. coli* carriage studies conducted. There have been various papers employing different typing methodologies and examining *E. coli* communities in a host of ways. For example, some studies focus on *E. coli* phylogroup distribution in river/sea water samples and often use the Clermont PCR method (Luo et al. 2010; Alwash and Al-Rafyai 2019). Additionally, several studies have examined the *E. coli* communities within farmed animals for consumption (Coura et al. 2015), domesticated household pets (Bourne et al. 2019) and wild animals such as bats (Nowak et al. 2017). Remarkably, carriage of pathogenic *E. coli* can be strikingly high in certain species such as bats and ownership of a dog can increase the likelihood of carrying a pathogenic *E. coli* in the owner's gut microbiome. Categorically, for the table below I will focus only on prior *E. coli* studies which have used faecal or wastewater treatment samples as a measure of human *E. coli* carriage.

There have been approximately 60 publications to date investigating phylotypes of *E. coli* in humans in different regions of the world and that of the surrounding environments (animals, foods, water etc). Our own interests as a research group have previously led us to compare *E. coli* phylotypes in the UK, Islamabad, Pakistan (Zahra et al. 2018), Silchar in Northeast India (Paul et al. 2020), and in Bangladesh (unpublished). We have confirmed on several occasions that there is a very real and distinct difference between *E. coli* phylogroup distribution in S. Asia (0-1% *E. coli* phylogroup B2 carriage) and South Wales (>30% B2 carriage). Several other studies have found distinct differences between different communities globally as well. In

2017 Stoppe *et al*, (Stoppe et al. 2017) used past literature from various countries to publish a worldwide statistical analysis comparing most of the previous publications on phylogroups, globally. Their conclusion was that there is **'no correlation between phylogroup distribution and geographic location, climate, living area, feeding habits, or date of collection'**. Although this paper serves as an excellent summation of prior phylogroup research into *E. coli,* we believe the rationale of this publication to be fundamentally flawed.

Most notably, the paper uses past studies from different countries as a representative of that country *E. coli* structure and composition. Only a small percentage of these studies would be appropriate to use as representatives of these given countries. As I will now demonstrate, most of these publications contain strain collections of *E. coli,* which are simply not appropriate to use as a measure of a regions 'normal' *E. coli* microbiome constituents.

The (**Table 5.1**) below is a modified version of Table 1 presented in Stoppe *et al* 2017 (Stoppe et al. 2017). This is the most comprehensive collection of past literature on indigenous *E. coli* carriage to date. I have amended the figure to focus more on the details of each study such as isolate number and the participants within the studies and omitted detail such as living area, feeding habits and geographic coordinates.

Table 5.1. Previous studies showing the distribution of the phylogenetic groups in different countries

Country/City	Year of sample collection	Type of sample collection	Number of participants in study	Number of *E. coli* isolates	Phylogenetic Group				References
					A	B1	B2	D	
Bay Reserve, Mali	1980-1990	Faecal	55	55 (One isolate per individual)	23.6%	58.2%	1.8%	16.4%	Duriez *et al*., 2001
Olib and Silba, Croatia	1980-1990	Faecal	57	57 isolates (One isolate per individual)	35.1%	31.6%	19.3%	14%	Duriez *et al*., 2001
Paris, France	1980-1990 & 1999-2001	Faecal	This data is an average of 3 data sets.	234 isolates across 3 studies	48.3%	9%	18.4%	24.4%	Duriez et al., 2001; Escobar-Páramo *et al*., 2004; Leflon-Guibout *et al*., 2008
Lahore, Pakistan	1984	Rectal swabs	22 Pakistani infants delivered at home	158 isolates	47%	18%	12%	23%	Nowrouzian *et al*., 2009
São Paulo, Brazil	1994-1997	Faecal	946 (children of outpatient clinics) & 890 (admitted to paediatric hospitals) children <10y/o with diarrhea	94 isolates	40.4%	8.5%	12.8%	38.2%	Carlos *et al*., 2010
Kyoto, Japan	~1995	Faecal	50 healthy adults	50 (One isolate per individual)	42%	32%	20%	6%	Kanamaru *et al*., 2006
Gothenburg, Sweden	1998-2001	Rectal swabs	70 infants from the ALLERGYFLORA study & 120 infants from the Flora study	482	30.5%	12.7%	42.5%	14.3%	Nowrouzian *et al*., 2005, 2006; Karami, 2007
Bogota, Colombia	1999-2001	Faecal	28 people (One isolate per individual)	28 (One isolate per individual)	57.1%	3.6%	25%	14.3%	Escobar-Páramo et al., 2004
Maroochydore, Australia	2009	Faecal swab	59 (30 healthy young adults and 29 elderly)	1541 (~28 per swab)	15.2%	10.6%	28.5%	45.7%	Vollmerhausen *et al*., 2011
Cotonou, Benin	1999-2001	Faecal	46 women	46 (One isolate per individual)	50%	32.6%	17.4%	0%	Escobar-Páramo *et al*., 2004
Alto de los Zarzos, Bolivia	1999	Faecal swab	72 inhabitants of remote Bolivian community	113 isolates	77%	10%	5%	8%	Pallechi *et al*., 2007
Brest, France	1999-2001	Faecal	21 University students	21 (One isolate per individual)	14.3%	23.8%	33.3%	28.6%	Escobar-Páramo *et al*., 2004
Tours, France	1999-2001	Faecal	24 pregnant women	24 (One isolate per individual)	25%	21%	29%	25%	Escobar-Páramo *et al*., 2004
Brittany, France	1999-2001	Faecal	25 bank workers and 25 pig farmers (One isolate per individual)	50 (One isolate per individual)	28%	26%	24%	22%	Escobar-Páramo *et al*., 2004
Tokyo, Japan	1999-2001 & 2011	Faecal	181 healthy subjects. Both men and women	181 (One isolate per individual)	16.6%	9.4%	50.3%	23.8%	Obata-Yasuoka *et al*., 2002; Harada et al.,
National Park, French Guiana	1999-2001	Faecal	93	93 (One isolate per individual)	63.4%	20.4%	3.2%	12.9%	Escobar-Páramo *et al*., 2004
Madrid, Spain	2001, and 2000-2005	Faecal	38 from 2001 with 3 months no hospital or antibiotic exposure. Remaining from 2000-2005 study	58 isolates	49.1%	19.3%	7%	24.6%	Machado *et al*., 2005; Valverde *et al*., 2009

Country/City	Year of sample collection	Type of sample collection	Number of participants in study	Number of *E. coli* isolates	Phylogenetic Group				References
					A	B1	B2	D	
Boise, United States	2002	Faecal	217 healthy volunteers	122 isolates	26.2%	13.9%	28.7%	31.1%	Hannah *et al*., 2009
Angaiza, Peru	2002	Faecal	89 villagers (0–59 y/o)	11 isolates	72%	17%	3%	8%	Bartoloni *et al*., 2009
Minneapolis, United States	2002	Faecal	Average of 147 healthy women across multiple studies	147 isolates	13.6%	15%	53.1%	18.4%	Zhang *et al*., 2002; Sannes et al., 2004; Johnson *et al*., 2005; Logue et al., 2012
Ann Arbor, United States	2002	Faecal	88 healthy Michigan women 18 to 39 (y/o)	88 isolates	20.5%	12.5%	47.7%	19.3%	Zhang *et al*., 2002
Oslo and Telemark, Norway	2002	Faecal	10 Healthy females (51–55 y/o)	20 isolates	30%	25%	15%	30%	Grude *et al*., 2007
Western France	2002-2004	Faecal	25 female nurses	25 (One isolate per individual)	48%	12%	20%	20%	Mereghetti *et al*., 2002
Barcelona, Spain	2003-2005	Faecal	39 healthy women (mean age 29.7)	120 isolates (3 per subject)	33%	19%	17%	31%	Moreno *et al*., 2009
Copenhagen, Denmark	2004, 2007, and 2009	Rectal swabs, faecal	This data is an average of 3 data sets. Dog owners, Inflammatory bowel disease patients and community dwelling humans	171 isolates from across 3 different studies	20.5%	22.2%	29.2%	28.1%	Damborg *et al*., 2009; Petersen et al., 2009; Jakobsen *et al.*, 2010
Villa Real, Portugal	2007-2008	Faecal	112 healthy children (1-14 y/o)	58 isolates	8.6%	10.3%	65.5%	15.5%	Silva *et al*., 2012
Munchengladbach, Germany	-	Faecal	37 from Max von Pettenkofer-Institut	37	43.3%	27%	16.2%	13.5%	Sorsa *et al*., 2007
Seoul, Korea	2008	Faecal	205 healthy adults from two Korean hospitals	205 (One isolate per individual)	32%	18%	22.9%	21%	Lee *et al*., 2010
Jeonnam Province, Korea	2008	Faecal	122 healthy adults	442	29.8%	34%	0%	36.2%	Unno *et al.,* 2009
Geneva and Ticino, Switzerland	2008	Faecal	10 healthy people	10 isolates	40%	10%	20%	30%	Grasselli *et al*., 2009
Fuzhou, China	2009	faecal	174 males and 151 females, with an age range from 18 to 75 years	325 isolates (one per person)	43.7%	23.4%	16%	16.9%	Li *et al*., 2010
Babylon, Iraq	2010	Rectal swab	10 from women	10 (One isolate per individual)	60%	30%	0%	10%	Abdul-Razzaq and Abdul-Lateef, 2011
Villamontes, Bolivia	2010	Rectal swabs	21 children	29 isolates	79.3%	0%	0%	20.7%	Riccobono *et al.,* 2012
Calgary, Canada	2011	Faecal swab	115 individuals	115 isolates (20 of which from healthy donors)	13.9%	12%	53.7%	20.4%	White *et al*., 2011
Beijing, China	-	-	-	-	12%	0%	47.8%	40.2%	Luo *et al*., 2011
Fargo, United States	2012	Faecal	179 were provided to us by J. Johnson, and 25 from human volunteers at North Dakota State University	204 isolates	16.2%	15.7%	53.9%	14.2%	Logue *et al*., 2012
São Paulo, Brazil	2017	Faecal	116 adults	116- 1 isolate per participant	48.3%	5.2%	16.4%	30.2%	Stoppe *et al*., 2017

Having an accurate picture and understanding of the *E. coli* structure of different global communities will help to shed light on our complex relationship with *E. coli*. The catalyst for Stoppe to publish this paper was likely a reaction to Duriez (Duriez et al. 2001a) suggesting that differences in *E. coli* populations globally could be due to dietary, climate or genetic factors. Unconvinced by the conclusions of Stoppe *et al*. 2017, we believe these factors could likely influence the structure of different *E. coli* populations; however, this is still to be proved. I will now briefly summarise some of the merits of the prior studies and using the first few publications listed in the above table I will now justify why I believe these studies are not suitable for statistical comparison.

5.1.1.1 "Worldwide Phylogenetic Group Patterns of *Escherichia coli* from Commensal Human and Wastewater Treatment Plant Isolates". Front Microbiol. 2017. Stoppe NC, Silva JS, Carlos C, Sato MIZ, Saraiva AM, Ottoboni LMM, Torres TT.

Beginning with Stoppe (Stoppe et al. 2017): very importantly, this paper demonstrated that sampling from wastewater treatment facilities could be used as a measure of *E. coli* carriage within a community. The authors collected *E. coli* isolates from human faecal samples and from wastewater treatment facilities. Phylogroups were determined using Clermont PCR and phylogroup distribution between faecal and wastewater was confirmed by a positive correlation (Mantel test, $r = 0.607$, $P = 0.046$). Critically, these sample numbers would ideally be larger- approximately 30 *E. coli* isolates must have been collected from each of the 5 sewage works. The authors show a positive correlation between faecal and sewage water phylogroup distribution in São Paulo but argue that for Spain, Portugal, Australia, and the United States there is a negative correlation. Importantly as pointed out by the authors – for

these other countries, the faecal and wastewater samples were not collected from the same areas and are not from the same time points- hence it is unsurprising that the samples are negatively correlated.

5.1.1.2 "Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants." J Med Microbiol. 2010 Bailey JK, Pinyon JL, Anantham S, Hall RM.

The phylogroup percentages for Bailey (Bailey et al. 2010) are summarised above as a means of two different studies, which used the same isolated collection. In the first publication, the authors focused on solely examining the phylogroup distribution within a cohort of 69 people and comparing the *E. coli* phylogroup distribution levels to other known similar studies. The second study used 20 people from the same 69 cohort and focused on what type of antibiotic resistance these isolates possessed. Merits of the second publication include the large sample sizes chosen. Although the group of 20 people was not very large, for each person 100 *E. coli* were picked and screened for resistance. The paper made interesting insights into potential explanations for transmission and carriage of different *E. coli* phylogroups. It noted that two of the subjects in the study who were a cohabiting couple shared an antibiotic susceptible strain. Additionally, it showed that two different subjects in the study who did not know one another - shared an isolate, which had identical characteristics and resistance genes. These same subjects happened to eat food from the same outlets. Most importantly, the sample collection period spanned 2 years and only 69 isolates were examined. This is far too small a sample and too broad a time scale to accurately describe the phylogroup composition of Sydney.

5.1.1.3 "Population structure and uropathogenic virulence-associated genes of faecal *Escherichia coli* from healthy young and elderly adults." J Med Microbiol. 2011 Vollmerhausen TL, Ramos NL, Gündoğdu A, Robinson W, Brauner A, Katouli M.

Unlike some of the other studies discussed, Vollmerhausen (Vollmerhausen et al. 2011) cannot be criticised for isolate sample numbers. The paper sought to understand the differences in phylogroup distribution between healthy young participants and older participants of both genders in the Gold Coast of Australia and generated interesting results. Elderly male subjects were statistically more likely to carry pathogenic *E. coli* types and possessed a greater *E. coli* diversity. Contrary to the initial hypothesis of the authors, young females in the study had a lower prevalence of uropathogenic virulence factors compared to the other groups and young men could potentially serve as a reservoir for uropathogen transmission via sexual intercourse. The study design achieved almost equal gender distribution between both groups and controlled for variables such as recent antibiotic usage and whether subjects had prior urinary tract infections. The phylogroup data presented above is a mean of both age groups and genders. The paper should be commended for being an interesting snapshot of Australian *E. coli* carriage in 2010 and also for how the paper teases apart *E. coli* characteristics between different demographics. One criticism would be a lack of antibiotic resistance screening which would have been interesting. And additionally, a criticism pointed out by the authors themselves would be that to have greater statistical confidence in their observations the number of participants should have been larger.

5.1.1.4 "Large-scale population structure of human commensal *Escherichia coli* isolates." Appl Environ Microbiol. 2004. Escobar-Páramo P, Grenet K, Le Menac'h A, Rode L, Salgado E, Amorin C, Gouriou S, Picard B, Rahimy MC, Andremont A, Denamur E, Ruimy R.

The work of Escobar-Páramo (Escobar-Páramo et al. 2004) makes for an interesting large study where the authors compared different *E. coli* phylogroup distributions globally. However, there are major flaws, which I will discuss. Notably, the samples being chosen as markers of a given community here are either too small or are too specific a group of people to use as a measure. For example, the phylogrouping data for Cotonou in Benin, Africa was solely comprised of 46 women and this data showed 0% phylogroup D. Having a large sample size of near equal gender distribution is essential because prior work such as Vollmerhausen (Vollmerhausen et al. 2011) have demonstrated that there can be differences in phylogroup distribution based on gender. The results for Benin are interesting, but it should not be regarded as representative of Cotonou. Additionally, despite 5, 10 isolates being chosen for each participant only 1 isolate is chosen at random for Clermont PCR typing. This creates a very small cohort of isolates for representing such a large region.

The data for Brittany in France presented here is an average of two data sets for Brittany. Within the study the authors collected 1 *E. coli* isolate from each of the 50 subjects. Of the subjects: 25 of which were bank and insurance workers (BIW) and the other 25 were pig farmers (PF). The B2 percentage for each group is quite substantially different. The B2 percentage for bankers was double that of pig farmers (16% PF compared to 32% BIW). These results are uniquely interesting alone but how can it be logical to then make an average of these two data sets as a

representative of Brittany. The original author has deliberately chosen these separate groups of people to study. Although admirable, the study tries to address too many questions without enough sample collection. The huge number of variables and small sample sizes makes it difficult to know whether these observable differences are real. The Brest and Tours data for France are again fraught with the same problems. These sample sizes are very small and contain very specific groups. Particularly Tours, in which 24 *E. coli* came from pregnant asymptomatic French women. The methods of the publication are unclear as to whether this control group was collected over a wide time span like the other isolates (1989 and 1997) and whether these control samples come from across France like the other isolates do. In which case these isolates may not even be representative of Tours. In the context of the original publication this control group of pregnant women made perfect sense, but Stoppe should not then try to use such a specific proportion of the population (i.e., only pregnant women) to represent all of Tours *E. coli*.

5.1.1.5 "Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure." Antimicrob Agents Chemother. 2007 Pallecchi L, Lucchetti C, Bartoloni A, Bartalesi F, Mantella A, Gamboa H, Carattoli A, Paradisi F, Rossolini GM.

The Pallecchi study (Pallecchi et al. 2007) is unique in its demographic. Two papers from this focused on the research question of what types of resistant strains of *E. coli* existed in a geographically isolated community in Bolivia which experienced minimal antibiotic exposure and living conditions of huts with poor hygiene and sanitation. Surprisingly, the authors showed that quite a lot of resistance mechanisms are present within the normal carriage of this isolated community. Explanations for why the resistance genes are so disseminated are unclear. However, although these

results are interesting, the phylogroup data presented here are only those of *E. coli*, which possessed antibiotic resistance. Initially, 108 members of the study were present but since 78 of which contained resistant isolates- these then became the isolates of focus and the ones, which were phylotyped. We know from our own research into measuring phylogroup distribution within a population that proportions of phylogroup distribution change significantly when looking with or without antibiotic selection (Toleman Group-unpublished data). Therefore these *E. coli* distributions should not be regarded as a representative of normal *E. coli* phylogroup distribution in Bolivia.

5.1.1.6 "Carriage of Antibiotic-Resistant Escherichia coli among Healthy Children and Home-Raised Chickens: A Household Study in a Resource-Limited Setting." Microbial drug resistance 2011 Eleonora Riccobono, Lucia Pallecchi, Antonia Mantella, Antonia Mantella, Gianni Maria Rossolini.

Riccobono (Riccobono et al. 2012) *et al.* sought to understand the antibiotic resistance carriage in young children in Peru and understand whether there was a relationship between home-reared chicken and a human-animal exchange of resistant *E. coli* occurring. The author's rationale for the study was resistance to tetracycline and quinolones in children is unusual since these antibiotics are not usually prescribed in childhood. Ultimately the paper found that although *E. coli* isolates were shared between households for children and separately for chickens, there was no link between children and chicken *E. coli* types. The paper fails to suggest that the resistance genes could have been acquired from a parent (particularly the mother- since early exposures at birth help to shape the microbiome) and one amendment could have been to screen the other members of the household also. Stoppe uses this reference as a measure of Bolivian *E. coli* phylogroup

distribution, yet the sample size is only 29 isolates from 21 children. As noted, *E. coli* phylogroup distribution changes across ages and genders within the same geographic region (Vollmerhausen et al. 2011). Hence, there is no rationale for using solely Bolivian children as a measure of a country's *E. coli* phylogroup distribution.

5.1.1.7 "*Escherichia coli* phylogenetic group determination and its application in the identification of the major animal source of faecal contamination." BMC Microbiol, 2010. Carlos, C., Pires, M.M., Stoppe, N.C. et al.

Carlos (Carlos et al. 2010) *et al.* 2010 compared phylogroup distribution between human and animal *E. coli* isolates (chicken, goat, cow, pig, and sheep). The aim of this study was to determine whether faecal contamination in São Paulo could be demonstrated by the presence of similar *E. coli* isolates in both human and animal hosts. Taken separately this paper shows an interesting application of *E. coli* typing for the betterment of human health in a cost-effective way. However, the data presented in Stoppe (Stoppe et al. 2017) *et al*, 2017 and in the above modified table is disingenuous. The 94 human isolates used in the study should not be represented as a picture of São Paulo's *E. coli* population because these isolates specifically come from a 4-year collection of faecal samples from children admitted to paediatric care below the age of 10 years old and presenting with diarrhoea. There is no way such a precise strain collection could be regarded as representing normal *E. coli* carriage.

5.1.1.8 "Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity." Appl Environ Microbiol. 2011 White AP, Sibley KA, Sibley CD, Wasmuth JD, Schaefer R, Surette MG, Edge TA, Neumann NF.

White (White et al. 2011) *et al* describes how 284 *E. coli* isolates were collected from human subjects (115 human-sourced isolates) and non-human sources such as water and animals. These isolates were sequenced and compared to understand the host specificity of different *E. coli* phylogroups. Individually this serves as an interesting publication but used in the context that Stoppe has presented I have my concerns. To present these phylogroup levels as representative of the Canadian population is not accurate. For example, the publication fails to give clear information on when these samples were collected or any characteristics of the subjects in the study such as age, gender etc. It is only stated that 20 of the 115 human isolates came from healthy donors. Within the study, it is unknown whether the other 95 isolates are clinically relevant to enteric symptoms or simply commensal. This data was withheld presumably to not bias the scientists investigating. However, this absence of essential information should have led Stoppe to discount this study from statistical comparison. If this collection of *E. coli* is indeed predominantly causing enteric symptoms, then to statistically compare a collection of pathogenic *E. coli* from one country to a small cohort of asymptomatic pregnant women's *E. coli* Escobar-Páramo (Escobar-Páramo et al. 2004) *et al.,* 2004 for example is obviously farcical.

5.1.1.9 "Phylogenetic groups and pathogenicity island markers in faecal *Escherichia coli* isolates from asymptomatic humans in China." Appl Environ Microbiol. 2010 Li B, Sun JY, Han LZ, Huang XH, Fu Q, Ni YX.

The study of Li (Li et al. 2010) *et al.* must be commended for its scale and clear study design. The authors recruited subjects who were already undergoing annual physical examinations in one hospital in Fuzhou. The study enrolled 325 subjects with a roughly equal split of male and female participants spanning a wide age range of 18-75 y/o. This study can be more accurately said to represent normal *E. coli* phylogroup carriage for the Fuzhou region compared to some of the other studies presented by Stoppe. Patients had no confirmed diagnosis of digestive tract diseases and *E. coli* were not selected for resistance profiles such as in other studies.

5.1.1.10 "Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations." Microbiology 2001. Duriez P, Clermont O, Bonacorsi S, Bingen E, Chaventré A, Elion J, Picard B, Denamur E.

The data presented for Croatia, part of the Paris data and Mali all come from the Duriez (Duriez et al. 2001b) study. The study was brilliant at its inception (2001) for comparing different populations globally and picking *E. coli* without selection and comparing phylogroup and virulence factor distribution. However, this study is not without its flaws. Most notably, is the absence of more substantial information pertaining to the subjects included in the study. The most concerning thing about the strain collection is that its sampling time is given over a range of a decade (1980-1990). With a collection time spanning a decade it is not suitable to say these values are truly representative of Croatia, Paris or Mali's *E. coli* populations.

5.1.1.11 "*Escherichia coli* shedding patterns in humans and dogs: insights into within-household transmission of phylotypes associated with urinary tract infections." Epidemiol Infect. 2009. Damborg P, Nielsen SS, Guardabassi L.

The phylogroup data presented for Copenhagen (Damborg et al. 2009) is very poorly chosen. It is an average of 3 very different data sets that are only unified by a similar geographic location. These data sets are years apart and are taken from very different studies. The first study was to gain insight into the possible transmission of pathogenic *E. coli* between dogs and their human owners. The second study Petersen (Petersen et al. 2009) *et al.* 2009 aimed to compare *E. coli* phylogroup distribution between patients with inflammatory bowel disease and healthy controls. And finally, the third study Jakobsen (Jakobsen et al. 2010) *et al,* 2010 compared a huge number of isolates across a wide variety of sources to look at a possible link between *E. coli* carriage, dietary habits, and urinary tract infection isolates. Stoppe would have been better suited to only use the *E. coli* data set from Jakobsen (Jakobsen et al. 2010) *et al,* 2010 as it contains a 109 faecal *E. coli* strain collection chosen without selection from healthy community dwelling citizens. This alone would provide a far more accurate representation of *E. coli* carriage in Denmark. However, by making an average using the other two data sets it undermines the reliability of the Jakobsen data set. As noted, there is no time continuity between these studies, so it is therefore not appropriate to collate them.

5.1.1.12 "Genetic relationship between *Escherichia coli* strains isolated from the intestinal flora and those responsible for infectious diseases among patients hospitalised in intensive care units." Hospital Inf. 2002. Mereghetti L, Tayoro J, Watt S, Lanotte P, Loulergue J, Perrotin D, Quentin R .

The isolates used to describe the *E. coli* composition of Western France (Mereghetti et al. 2002) come from only 25 isolates which were collected from female nurses between 2002-2004 as a control for a study looking to compare normal intestinal *E. coli* with *E. coli* isolated from ICU patients with invasive diseases. As shown in Vollmerhausen (Vollmerhausen et al. 2011) *et al,* 2011 differences in *E. coli* structure exist between different genders within the same community. Therefore, solely using 24 women as a representative for all Western France is a clear oversimplification. Additionally, the sample size is so small that Group B2 is the dominant phylogroup because there is one more isolate compared to group A and D. It could well be the case that group B2 is most prevalent, but 25 isolates are a tiny collection to base Western France upon.

5.1.2 Concluding remarks regarding past literature

As noted in the concluding remarks of Duriez (Duriez et al. 2001b) *et al,* 2001: "Studies on a large number of commensal strains including precise geographic, socioeconomic and medical data are needed to provide further insight into the emergence of virulent *E. coli* clones". I believe Stoppe *et al,* 2017 admirably sought to do this and in doing so did very well to create such a comprehensive record of the past literature. I hope through my dissection of each of these prior studies I have demonstrated unequivocally why these case studies are unsuitable for statistical comparison because of how significantly different each type of *E. coli* collection is

(e.g long sampling times, small number of isolates, bias for gender/age in subjects of study, bias of health of subjects of study).

5.1.3 UK sepsis information leading to our hypothesis

The UK ESBL study highlighted in Chapter 3 of this thesis strengthened our resolve for using sewage as a measure of pathogenic *E. coli* phylotypes that are carried in the community. This is because the study demonstrated a direct link between *E. coli* phylotypes causing bacteraemia in hospital patients to the carriage of the same phylotypes in human faeces and those found in municipal sewage Day, *et al.* 2019 (Day et al. 2019). At the same time, we became interested in several facts concerning *E. coli* sepsis in the UK. Firstly, it is clear from the mandatory surveillance of UK *E. coli* bacteraemia that the *E. coli* sepsis rate has steadily been rising year by year and has more than doubled over the last decade. Secondly, the sepsis rates in England and especially London compared to the devolved nations are substantially lower (**Figure 5.1**). Based on our prior observations that the phylotypes of *E. coli* that are carried in the communities in different nations could be substantially different, we hypothesised that "**The differences in sepsis rates could simply reflect differences in carriage of pathogenic phylotypes within the UK."**

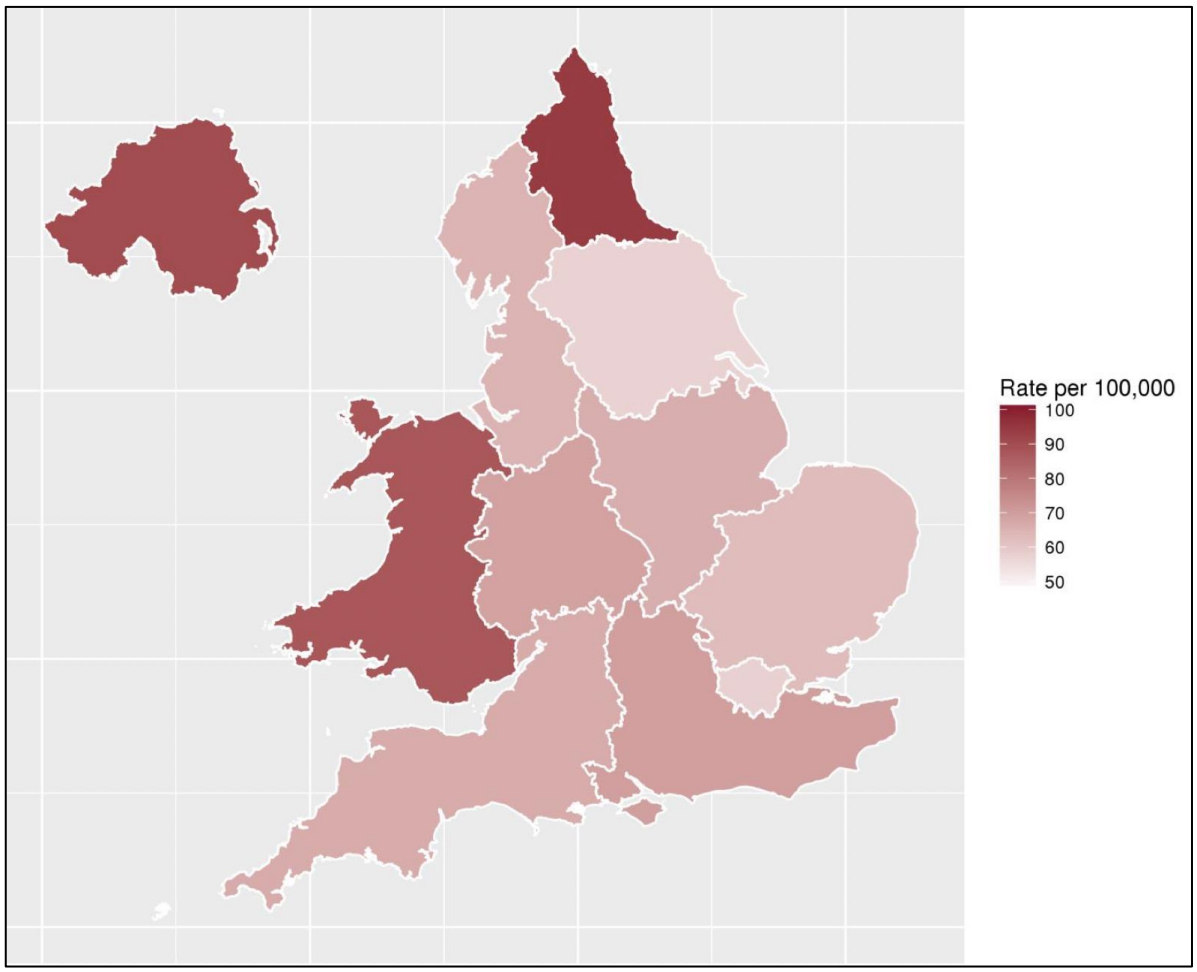


Figure 5.1. Geographical distribution of pathogenic *E. coli* bacteraemia per 100,000 population (England, Wales and Northern Ireland): 2017
Screenshot from Health Protection Report Volume 12 Number 22 22 June 2018.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/718820/hpr2218_ecoli.pdf

5.1.4 Study design

To understand the pervasiveness of pathogenic *E. coli* types in the UK community we sought to assay their presence in human sewage without any antibiotic selection. We hypothesised that the huge volumes and homogenation within the sewage network upon entry to each waste-water facility act to normalise many tens of thousands of human gut samples into a single sample, representing tens of thousands of individuals. Furthermore, to understand differences in *E. coli* sepsis rates between the UK nations we focussed on the consistent higher rates of sepsis observed in Wales as compared to England. We therefore sampled sewage on entry

to waste-water treatment facilities at either end of the M4 motorway corridor which links Cardiff to London.

The utility of using sewage as a surveillance tool has historical links starting with William Budd and John Snow in the 19th century and in the 1960's with polio detection by Swedish scientists. More recently, COVID-19 outbreaks have been accurately identified through sewage surveillance during the pandemic. Moreover, the Stoppe study in Brazil demonstrated paralleled *E. coli* phylotype distribution between rectal and sewage samples. This has also been shown for the UK, where ESBL *E. coli* MLST prevalence mirrored that of >20,000 rectal swabs collected in five regions across the UK.

Community carriage of *E. coli* was ascertained by sampling untreated sewage on entry to lagoons at each waste-water facility. Individual samples of untreated sewage (c. 100 mL) were collected at 6 waste-water treatment stations situated along the M4 Motorway between Cardiff and London from 19th to 26th September 2019. Three of these sites were towards the west end of the M4 within 40 miles of Cardiff (Cardiff, Newport and Bristol) and three were at the East end of the M4 motorway within 40 miles of London (Longreach, Marlow and Reading). To ensure accuracy between and within samples, the Newport sample was assayed three times, and a second Newport sample collected a month later (October 31st, 2019) was assayed a further two times (5 replicates in total). As a comparison, and because of our prior knowledge of very low carriage rates of pathogenic ExPEC *E. coli* in South Asia, we also sampled sewage at approximately the same time (October 21st) in Dhaka, Bangladesh.

5.2 Data collection and methodology rationale

5.2.1 *E. coli* colony isolation

Each sample was thoroughly mixed before dilution and plating on Brilliance™ UTI Clarity Agar (Oxoid) plates, without antibiotics. These growth plates give a presumptive species ID based on colony colour. The plates were incubated for 18 hours at 37°C before picking individual colonies. Samples were diluted (typically 2,000 x) so that only a total of five to 10 red colonies (presumptive *E. coli*) were visible on each plate. Each whole sample (c. 100mL) was then spun in a centrifuge to collect bacteria and the pellet stored by freezing in 3X cryotubes with storage buffer (2 mL) for future analysis and comparison. Operator bias in choice of colonies was removed by picking all red colonies on each plate, irrespective of size, morphology or depth of colour. From each well-mixed sample, a total of c. 150 *E. coli* colonies were isolated after dilution and plating on Brilliance ™ UTI Clarity™ Agar (Oxoid) giving a presumptive *E. coli* species identification (typically from 20-25 plates). The presumptive ID was then confirmed by MALDI-TOF and the isolates that were not *E. coli* (often various *Aeromonas* or *Citrobacter* spp.) discarded.

5.2.2 *E. coli* phylotype B2/D genome sequencing/Bioinformatics and phylogenetics

Phylotype was determined for all isolates by Clermont phylotype PCR and Sequence types were initially determined by the Doumith PCR. However, because the Doumith PCR gave misleading results for some isolates (**Table 5.2**) all UK phylotype B2 and D isolates and Bangladesh phylotype D isolates (no B2 found) were further submitted to Whole Genome Sequencing by Illumina Miseq technology. From the sequences the MLST types were determined as well as the complement of antibiotic resistance encoding genes and virulence genes.

Table 5.1.Example of *E. coli* isolates misidentified by Doumith ST PCR

Isolate ID	Origin	Phylotype based on PCR	Virulence ST
b8	Bristol	B2 ST95	ST12
b16	Bristol	B2 ST95	ST929
b77	Bristol	B2 ST95	ST95
b78	Bristol	B2 ST95	ST95
b82	Bristol	B2 ST95	ST547
b92	Bristol	B2 ST95	ST95
b117	Bristol	B2 ST95	ST131
b120	Bristol	B2 ST95	ST131
b2-6	Bristol	B2 ST95	ST95
b2-12	Bristol	B2 ST95	ST127
b2-15	Bristol	B2 ST95	ST131
c27	Cardiff	B2 ST95	ST95
L7	Longreach	B2 ST95	ST 12
L125	Longreach	B2 ST95	ST12
M1a	Marlow	B2 ST95	ST 1231
p105	Ponthir	B2 ST69	ST95
p48	Ponthir	B2 ST95	ST95
p55	Ponthir	B2 ST95	ST 536

Some ST (especially ST95) were commonly misidentified by the Doumith ST PCR in our study. The table gives examples of isolates that were identified as ST95 by PCR but proved to be other ST by sequence analysis. Yellow boxes represent isolates, which were successfully sequenced (note: b2-6 is not yellow because this isolate was lost).

5.2.3 Statistical analysis

We used the same statistical approach to sample collection as discussed by Lautenbach *et al* 2008. Specifically, with the infinite number associated with the *E. coli* count in a sewage works the multivariate hypergeometric distribution is well approximated by the multinomial distribution. That is, the probability of selecting at least one colony of a specific type from **n** sampled colonies, with a proportion **p** of organisms of this type, is computed using the multinomial distribution ie:

$$1 -(1 -p)^{n}.$$

For our current studies we have chosen to take sample numbers of c. 100 *E. coli*. Thus, the probability of identifying an *E. coli* ST that represents 5% or 3% prevalence would be 99% and 95%, respectively. Differences in the prevalence of *E. coli*

belonging to phylotype B2 and D between sites West and East of the M4 and between England and Wales were assessed using the C^2 Pearson statistical test. This test was also used to determine the within and between assay accuracies of our approach.

5.3 Results

5.3.1 Totals of *E. coli* isolated from each sample site

Total confirmed *E. coli* from each site were: Cardiff, 98; Newport, 116; Bristol, 112; Reading, 121; Marlow, 91; Longreach, 96; Total 634). Circa 100 *E. coli* isolates were also collected from duplicate repeats of the original Newport frozen sample (123 & 90 isolates, September 19th) as well as duplicate isolations from a subsequent sample from the Newport waste-water treatment facility (99 & 104 isolates October 31st) with a total of an additional 416 *E. coli* isolates (5 separate assays on two samples collected 1 month apart from Newport site 532 total *E. coli* tested by phylogenetic PCR). These were isolated and confirmed as *E. coli* in an identical fashion. All 1050 UK isolates were further phylotyped by Clermount phylotype PCR into one of 5 *E. coli* phylotypes (A, B1, B2, C, and D). The initial 634 *E. coli* were also assayed for antibiotic resistance by disc diffusion using CLSI guidelines. *E. coli* from Bangladesh sewage were isolated in an identical fashion from pooled sewage isolated from 9 sites around the city. A total of 111 *E. coli* isolates were confirmed by MALDI-TOF and phylotyped.

5.3.2 *E. coli* phylotype proportions in the UK and Bangladesh

PCR-based phylotype analysis of *E. coli* isolated from all UK initial samples (634) indicated that proportions of phylotypes A-D varied in overall prevalence: A 30.7%, (196); B1 19.9%, (124); B2 21.2%, (134); C 11.9%, (77); D 14.2%, (90). The

phylotype of 12 isolates could not be determined. The phylotype proportions from Bangladesh (111) were A (30.6%), B1 (64.9%), B2 (0%) and D (8.1%), (**Table 5.2**). Phylotypes E and F were not represented in either UK or Bangladesh collections and phylotype C was not represented in Bangladesh.

Table 5.2. *E. coli* phylotype proportions at each sample site

	Non-pathogen			Pathogen		% Non-pathogen	% Pathogen
Site	**A (%)**	**B1 (%)**	**C (%)**	**B2 (%)**	**D (%)**	**(A + B1 + C)**	**(B2 + D)**
Cardiff	21.4	22.5	14.3	25.5	16.3	58.2	41.8
Newport	28.5	15.5	10.3	32.8	12.9	54.3	45.7
Bristol	22.3	17.0	21.4	23.2	16.1	60.7	39.3
Total West	**24.1**	**18.3**	**15.3**	**27.2**	**15.1**	**57.7**	**42.3**
Longreach	47.9	12.5	14.6	17.7	7.3	75.0	25.0
Marlow	23.1	34.1	8.8	14.3	19.8	66.0	34.1
Reading	42.2	18.2	13.2	12.4	14.1	73.6	26.5
Total East	**37.7**	**25.6**	**12.2**	**14.8**	**13.7**	**71.5**	**28.5**
UK in total	30.7	19.9	11.9	21.2	14.2	**64.6**	**35.4**
Bangladesh	30.6	64.9	-	0	8.1	**91.9**	**8.1**

5.3.3 Statistical analysis of replicate samples from the Newport waste-water site

We used the Chi-squared Pearson test and R-statistics software to assess the veracity of our data. Using the triplicate sample collections from the Ponthir site in Newport, September 2019, and the duplicate samples from October 2019. The Chi-squared analysis indicates very low P values within the samples indicating the very low probability of getting the different phylogroup prevalence in each collection by chance. The analysis between the five samples taken September/October indicates that there is no difference between these samples with a P value of 0.562 (**Figure 5.2**).

5.3.4 Statistical analysis of samples from different UK locations

We used the same methodology to determine that samples from each site were statistically different both within the sample and between the different locations and are unlikely to have arisen by chance (**Figure 5.2 & 5.3**).

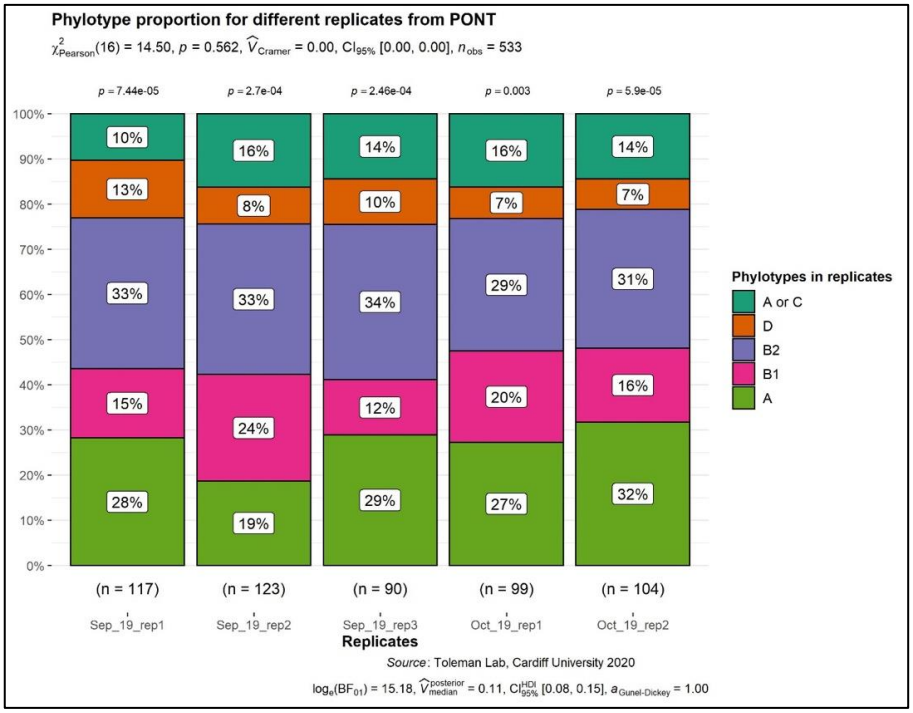


Figure 5.2. Statistical analysis of multiple samples from the Newport wastewater treatment site

P value within samples (<0.1) indicate that phylotype proportions in each sample have not arisen by chance. Whereas the p-value between replicates (0.562) indicates that there is no statistical difference between replicate samples.

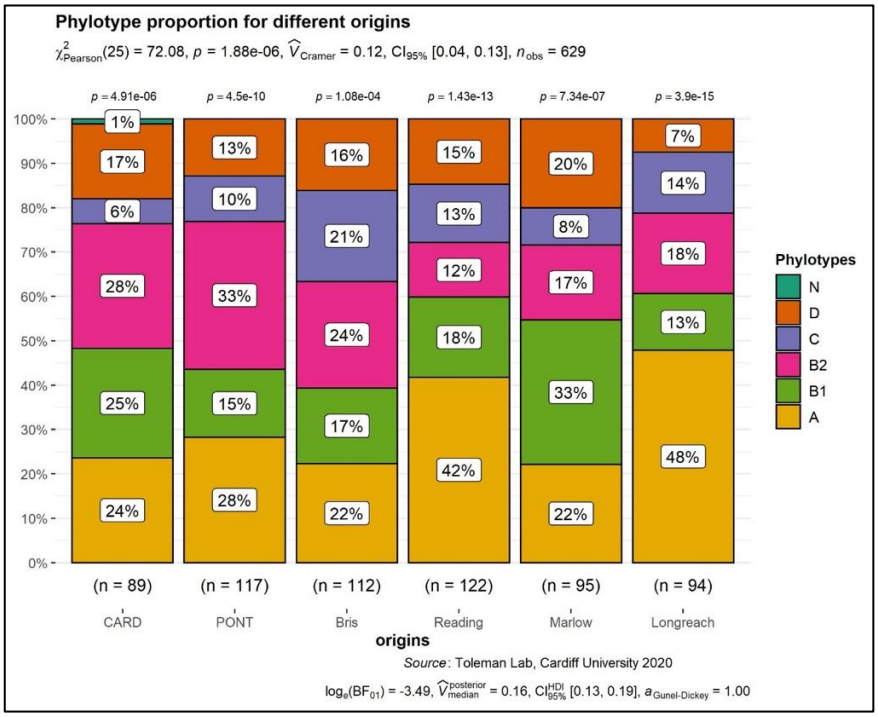


Figure 5.3. Statistical analysis of samples collected from different UK sites

P value within samples (<0.1) indicate that phylotype proportions in each sample have not arisen by chance. In comparison to (Figure 5.2) above, the p values between individual waste-water locations are statistically different p (<0.1).

5.3.5 *E. coli* Phylotype proportions within the UK.

Within the UK, each location had unique phylotype distributions. In particular, the pathogenic B2 phylotype varied at each site and ranged between 12.4% (Reading) - 32.8% (Newport) of all *E. coli* at each location (**Table 5.2 & Figure 5.3**). There was a statistically relevant difference between the proportion of pathogenic (B2 & D) vs non-pathogenic phylotypes (A, B1, C) collected at sites located at east and westerly ends of the M4 corridor: M4 east: 42.3% vs 57.7% and M4 west 28.5%.vs 71.5%, respectively (total *E. coli* east and west, 308 vs 326, respectively). This difference was almost entirely due to a statistically relevant difference in the prevalence of the B2 phylotype: M4 east, 27.2% vs M4 west, 14.8% (**Figure 5.4**) and was also statistically relevant for higher carriage rate of pathogenic *E. coli* B2 types between Wales and England (**Table 5.2, Figure 5.4 & 5.5**).

5.3.6 Differences between carriage of pathogenic *E. coli* in the UK and Bangladesh

Between countries differences in phylotype B2 carriage were stark with the prevalence of pathogenic phylotypes much higher in the UK compared to Bangladesh (B2: 21.2% vs 0%, D: 14.2% vs 8.1%), (**Table 5.2**).

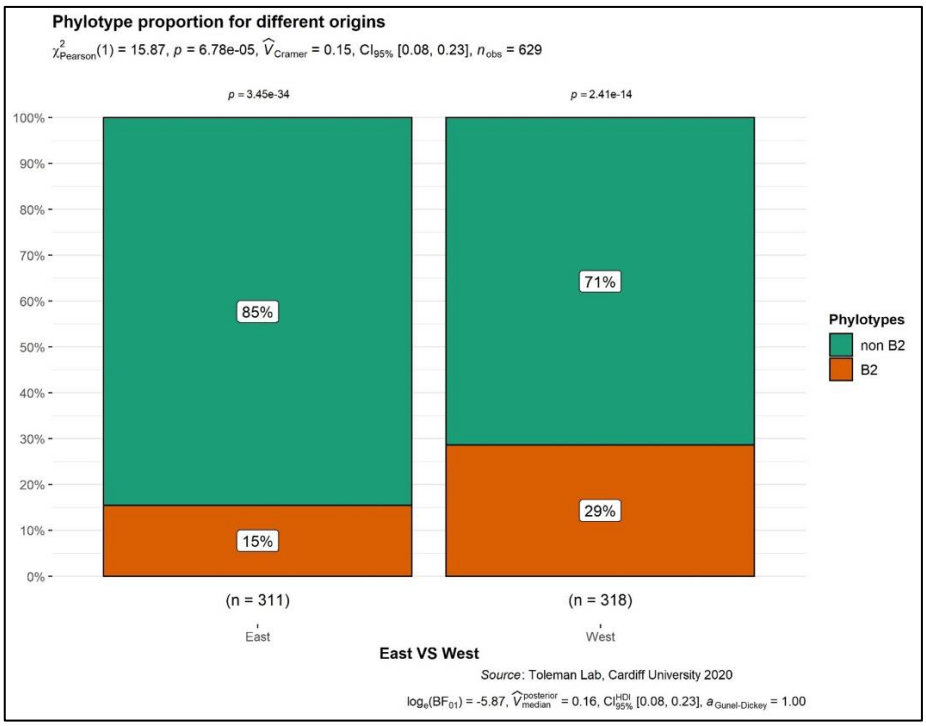


Figure 5.4. Statistical analysis of difference between pathogenic *E. coli* phylotypes East vs West

P value of (<0.1) indicates a statistical difference between carriage of phylotype B2 *E. coli* between the East and West of the UK.

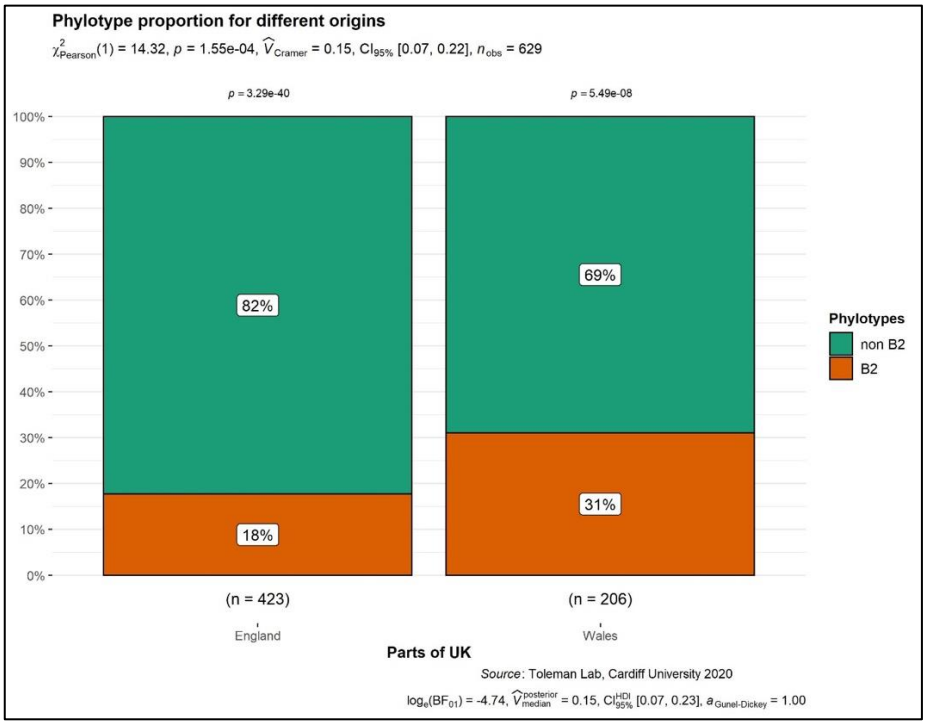


Figure 5.5. Statistical analysis of difference of carriage of pathogenic B2 phylotype *E. coli* between Wales and England

A p-value (<0.1) indicates a statistical difference in carriage of B2 *E. coli* between Wales and England.

5.3.7 Multi Locus Sequence type analysis of B2 and D pathogenic *E. coli* phylotypes

All B2 and D *E. coli* were whole genome sequenced using Illumina Miseq technology. MLST analysis of sequenced UK *E. coli* phylotype B2 and D strains (223) indicated that our collection was dominated by multiple examples of only a few individual MLST types (Figure 5.6). These were mostly of phylotype B2 and to a lesser extent phylotype D.

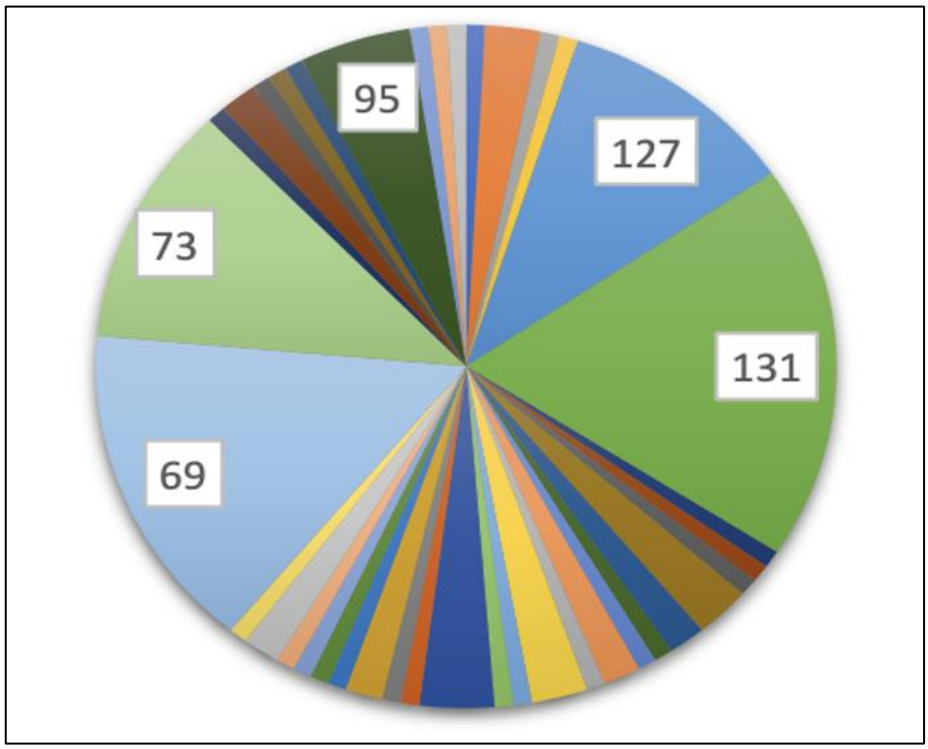


Figure 5.6. The UK *E. coli* carriage isolates belonging to phylotypes B2 and D are dominated by only a few MLST
Pie chart of 234 *E. coli* phylotype B2 and D isolates. The most prevalent being ST131, ST73, ST127, ST95, ST404 belonging to phylotype B2 and ST69 belonging to phylotype D. These accounted for 45% (100/234) of all pathogenic MLST.

The most common MLST belonging to the B2 phylogenetic group included ST131, ST127, ST73, ST95, ST404, ST12, and ST141. These represented 63% (84/134) of all B2 isolates in our collection (**Figure 5.7**). The remaining B2 isolates were mostly singleton's (27%, 36/134). The phylotype D strains were similarly dominated by

ST69 accounting for 28% (25/89) of all phylotype D strains and to a lesser extent ST38 (7%, 6/89) and ST349 (7%, 6/89) (**Figure 5.7**).

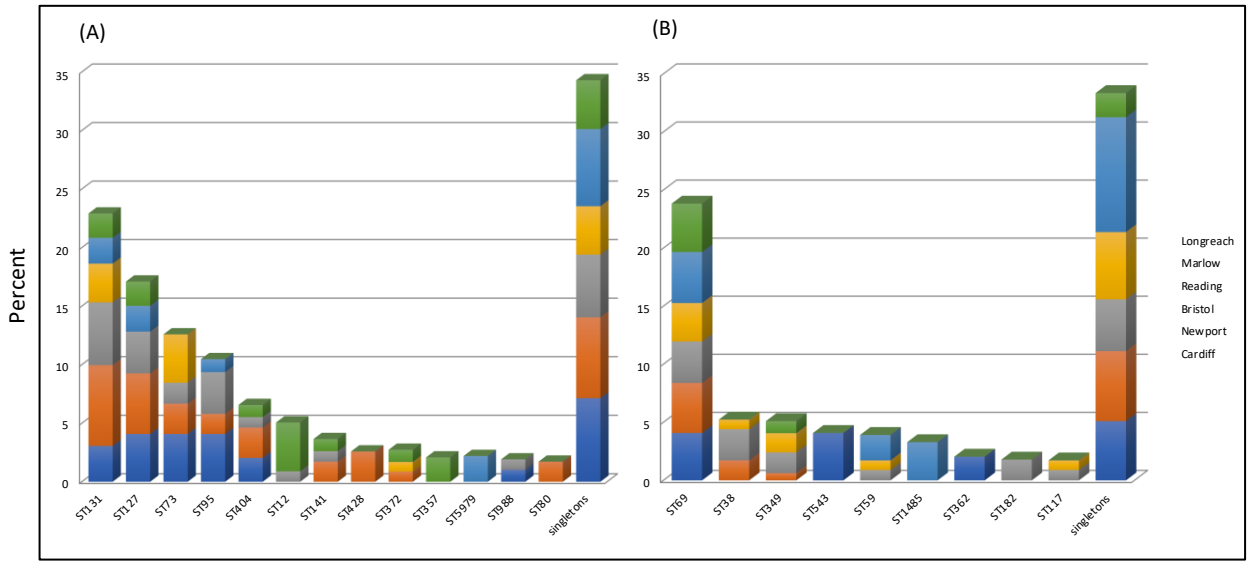


Figure 5.7. Displays the prevalence of each *E. coli* MLST group at each location belonging to: (A) Phylotype B2; (B) phylotype D as a percentage of all phylotype B2 and D isolates, respectively

Bangladesh phylotype D strains all belonged to novel MLST groups not described before. The six most prevalent pathogenic MLST found in our study ST131, ST69, ST127, ST73, ST95, ST404 together accounted for 45% (100/234) of all pathogenic phylotype isolates, and over 15% (100/645) of all UK *E. coli*. These were found at all westerly sites and all Welsh sites but not at all easterly locations (**Figure 5.7**). ST131 and ST69 were the most prevalent pathogenic strains and found at all locations (**Figure 5.7**). In addition to overall B2 prevalence being higher at westerly sites, the abundance and diversity of B2 MLST types (reflected in the number of common MLST types and singletons) was higher in Wales as compared to the London region (**Figure 5.7 & 5.8**). However, this was not seen for ST69 which was found at a near-identical prevalence at all sites (**Figure 5.8**).

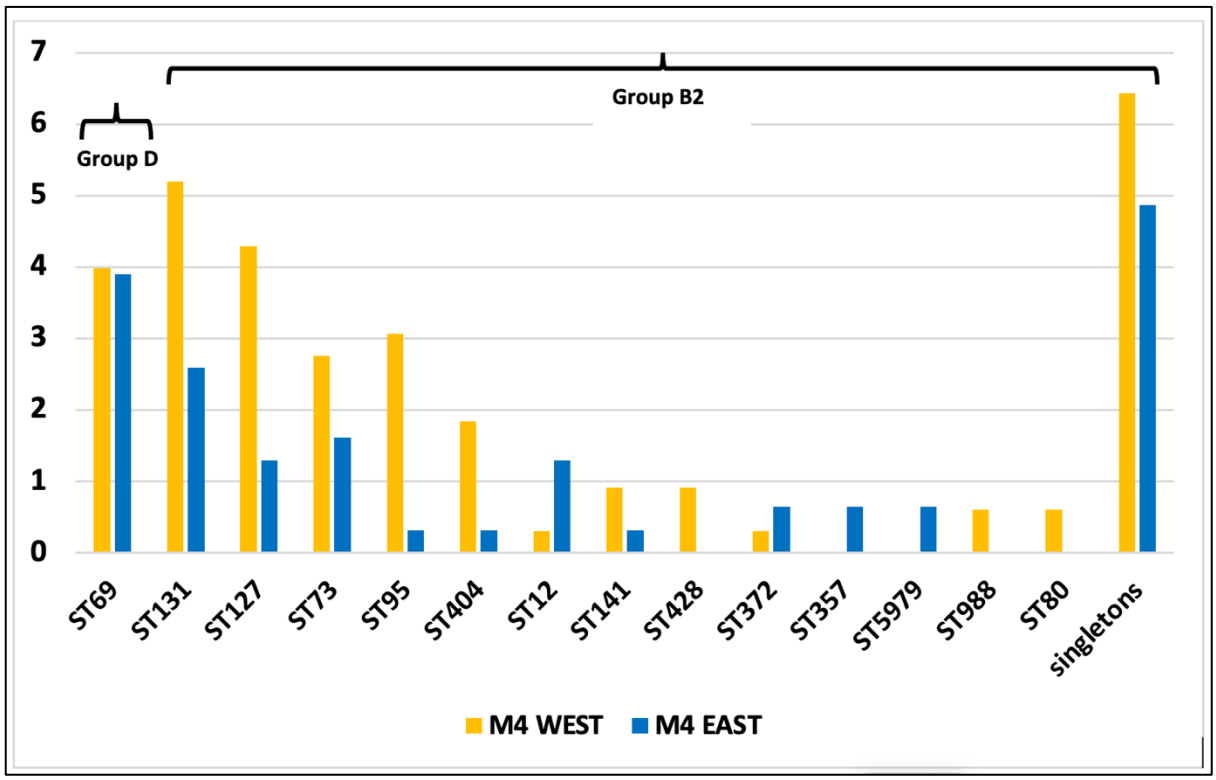


Figure 5.8. Prevalence of the most common pathogenic MLST in the UK. The Bar chart gives the % of each pathogenic MLST as a proportion of all B2 and D isolates in the collection

The difference in phylotype B2 carriage between sites in the West and sites in the East of the M4 corridor was mostly due to a higher carriage rate of ST131, ST127, ST73, ST95 and ST404 (**Figure 5.8**) in the west. The combined information on sepsis rates and carriage rates of pathogenic *E. coli* phylotypes and multi-locus sequence types are summarised in (**Figure 5.9**).

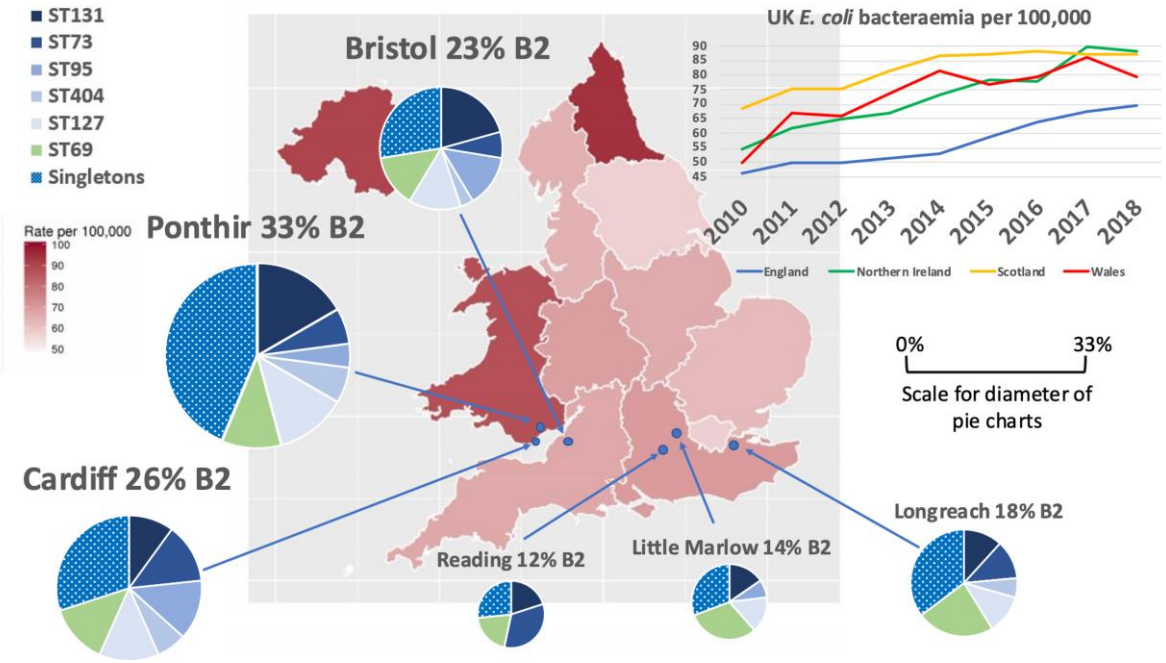
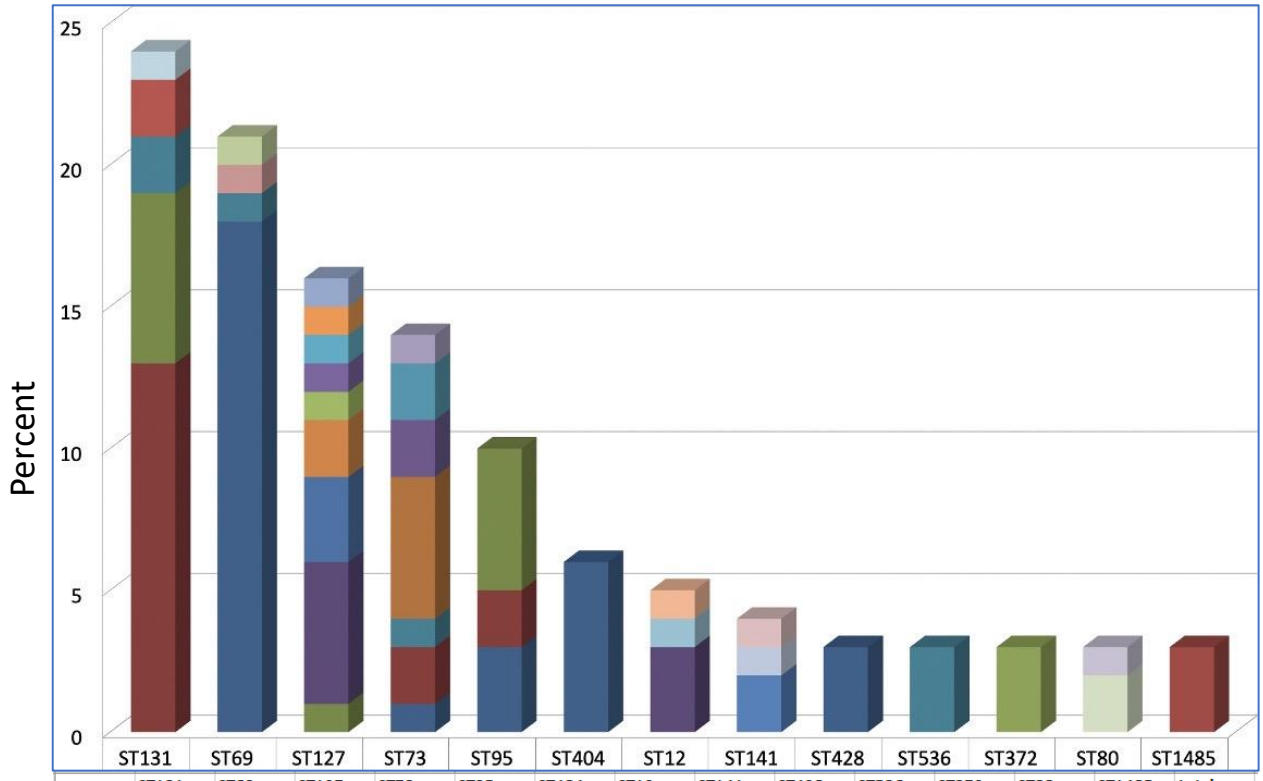


Figure 5.9. Summary figure of UK *E. coli* carriage Phylotypes and Sequence types

The figure is adapted from a map showing distribution of *E. coli* bacteraemia across the UK in 2017 (Health Protection Report, volume 12 Number 22, 22 June 2018). Pie charts represent the 6 most prevalent pathogenic *E. coli* MLST, B2 (ST131, ST127, ST73, ST95, ST404) and D (ST69), at each site together with singletons (B2 MLST mostly found as single copies). The diameters of the pie charts are proportionately sized according to percentage prevalence of phylotype B2 *E. coli* compared to other phylogroups at each site. The graph in the top right corner shows UK *E. coli* bacteraemia rates per 100,000 people in England, as compared to devolved nations.

5.3.8 *FimH* analysis.

We identified the various *fimH* alleles in the most common MLST groups among the pathogenic phylotypes using the *fimH* typing tool at the Centre for Genomic Epidemiology website. This was primarily to further define the isolates belonging to ST131, since it is known that quinolone and cephalosporin resistance is closely linked to isolates that possess the *fimH30* allele but also to see if successful isolates share common alleles known to promote strong binding to epithelial cells. The most common allele found was *fimH30* which was found in 13 ST131 isolates and in ST73 (2) and ST95 (2) isolates (**Figure 5.10**).



	ST131	ST69	ST127	ST73	ST95	ST404	ST12	ST141	ST428	ST536	ST372	ST80	ST1485	total
fimH22	1													1
fimH120												1		1
fimH1												2		2
fimH489								1						1
fimH130								1						1
fimH106							1							1
fimH6							1							1
fimH504				1										1
fimH483		1												1
fimH47		1												1
fimH341			1											1
fimH136			1											1
fimH126			1											1
fimH121			1											1
fimH100			1											1
fimH?	2													2
fimH5								2						2
fimH181			2											2
fimH103				2										2
fimH12				2										2
fimH9											3			3
fimH58													3	3
fimH54			3											3
fimH10				5										5
fimH0	2	1		1						3				7
fimH2			5				3							8
fimH41	6		1		5									12
fimH30	13			2	2									17
fimH27		18		1	3	6			3					31
total	**24**	**21**	**16**	**14**	**10**		**5**	**4**	**3**	**3**	**3**	**3**	**3**	

Figure 5.10. *fim*H alleles found in prevalent pathogenic UK multi-locus sequence types

Various *fimH* alleles were assigned to different MLST groups using the Center for Genomic Epidemiology website.

Other common *fimH* alleles were *fimH27* which was found mostly in ST69 (18 isolates) but also in ST404 (6), ST428 (3) ST95 (3) and ST73 (1) isolates and *fimH41* found in 6 ST131 isolates and in ST95 (5) and ST127 (1). ST127 isolates

harboured a diverse range of different *fimH* alleles including 2, 41, 54, 100, 121 126, 136, 181 and 341.

5.3.9 Antibiotic resistance among carriage *E. coli*

The collection of many isolates without any antibiotic bias is an ideal opportunity to assess the carriage of antibiotic resistance in UK *E. coli*. Antibiotic disc diffusion was performed on all isolates from each site using 14 different antibiotics and antibiotic/inhibitor combinations.

Table 5.3. Antibiotic resistance rates in all 634 isolates collected without antibiotic selection

Antibiotic	% resistant by Phylogroup								
	A	B1	B2	C	D	ALL	ESPAUR 2019	SCOTLAND 2019	WALES 2017
trimethoprim	5.6	16.9	11.1	16.1	12.1	**14.1**		37	42
sulphomethoxazole/trimethoprim	7.1	14.5	16.3	23	22	**15**			
mecillinam	7.6	4.8	8.1	3.4	5.5	**6.3**			
cephalexin	9.6	16.9	11.1	16.1	12.1	**15.4**			
cefuroxime	9 *	15*	22*	15*	11*	**98(i)**			
cefotaxime	4.1	5.6	19.3	11.5	4.4	**8.7**	22		
ceftazidime	14.2	25.8	27.4	17.2	25.2	**21**	22	9	12.9
amoxycillin	83.2	89.5	90.4	82.8	85.7	**86**			
amoxycillin/clavulanic acid	14.2	21.8	28.9	16.1	8.8	**18**	44	53	37
piperacillin/tazobactam	10.2	12.9	14.8	14.9	8.8	**12**		8.5	12.9
ciprofloxacin	5.6	14.5	13.3	25.3	9.9	**12**	20	19	20.3
gentamycin	9.6	3.2	9.6	13.8	7.7	**8.6**		11	11
nitrofurantoin	1	6.4	4.4	1.2	3.2	**3.1**			11
fosfomycin	1	12.1	2.2	1.1	1.1	**3.4**			
Average antibiotic resistance	**6%**	**18.6%**	**19.9%**	**18.3%**	**15.54%**				

* intermediate resistance to cefuroxime was high in the collection (A: 98.5%, B1: 100%,B2: 97%, C: 95%, D: 95%).

Isolates were tested for antibiotic resistance by disc diffusion using CLSI methodology. Resistance rates gleaned from various studies are also given as comparators.

Results indicate that overall resistance was lowest in phylogroup A and highest in the phylotype B2 group. Antibiotic resistance was particularly high for amoxycillin (86%) and 98% of isolates were of intermediate resistance to cefuroxime. Fosfomycin and nitrofurantoin resistance was higher in group B1 isolates but generally low across the collection. 3rd generation cephalosporin resistance was also high in the collection averaging 21% for ceftazidime and 9% for cefotaxime with the

highest rates in the B2 isolates (27.4% ceftazidime resistance and 19.3% cefotaxime resistance, (**Table 5.3**). Ciprofloxacin resistance was highest in phylotype C (25.3%), which also had the highest rates for sulphonamide resistance (23%,**Table 5.3**). There were minor variations generally in resistance rates by location (**Figures 5.11 and 5.12**) and with rates generally following the overall resistance by phylotype analysis. However, there were sites that had higher resistance rates for some antibiotics: for example, Longreach had high resistance rates for cefotaxime among B2 isolates and ceftazidime resistance was higher in the Bristol area than other sites . There were also sites where resistance for some antibiotics was absent: for example, Fosfomycin resistance was absent from isolates collected from Reading, Marlow and Newport sites. Nitrofurantoin resistance was absent from the Newport and Marlow sites (**Figures 5.11 and 5.12**). Gentamicin resistance was also absent from the Longreach site.

5.3.10 Genetic basis of antibiotic resistance

Sequence analysis of the prevalent B2 and D MLST types revealed the genetic basis of several antibiotic-resistance phenotypes. ESBL genes were common in ST131 strains including 4 strains with *bla*$_{\text{CTX-M-27}}$, 3 strains with *bla*$_{\text{CTX-M-15}}$ and one strain with *bla*$_{\text{CTX-M-10}}$ but were not found in other sequence types (**Figures 5.13 and 5.14**). The β-lactamase gene *bla*$_{\text{TEM}}$ and several variants were commonly found in many of the prevalent pathogenic MLST and especially in ST131 and ST69 isolates (**Figures 5.13 and 5.14**). It is unclear from the literature whether any of the *bla*$_{\text{TEM}}$ variants in our collection are ESBLs (BLDB website). There was some geographic bias associated with *bla*$_{\text{TEM}}$ in ST73 isolates which were found in Newport and Reading sites only. Also, it was found in the collection more often at western than Eastern sites. The most prevalent MLST, ST131 and ST69 harboured more resistance genes

than all other ST, likely including a class 1 integron harbouring *dfrA17* or *dfrA12* genes conferring trimethoprim resistance and aminoglycoside resistance via *aadA5* or *aadA2* as well as sulphonamide resistance via *sul1*.

Figure 5.11. Percent resistance by phylotype for *E. coli* collected from each sample site

Figure 5.12 Percent resistance by phylotype for *E. coli* collected from each sample site

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	isolate ID	origin	date	SEQUENCE	Phylotype	Virulence ST																
2	b85	Bris	19-Sep	SEQ 4/2020	B2 ST131	ST131	fimH30	blaEC-5	blaTEM-1		dfrA12	aadA2	aac(3)-IId			tet(A)	sul1	mph(A)	catA1			
3	b96	Bris	19-Sep	SEQ 5/2020	B2 ST131	ST131	fimH30	blaEC-5	blaCTX-M-10		dfrA17	aadA5	aph(3'')-Ib	aph(6)-Id		tet(A)	sul1, 2	mph(A)			S83L gyrA	
4	b117	Bris	19-Sep	SEQ 5/2020	B2 ST95	ST131	fimH0	blaEC-5	blaTEM-1													
5	b120	Bris	19-Sep	SEQ 5/2020	B2 ST95	ST131	fimH30	blaEC-5												D87NgyrA	S83L gyrA	
6	b2-15	Bris	19-Sep	SEQ 5/2020	B2 ST95	ST131	fimH30	blaEC-5	blaCTX-M-27		dfrA17	aadA5	aph(3'')-Ib	aph(6)-Id		tet(A)	sul1, 2				S83L gyrA	
7	c4	CARD	24-Sep	SEQ 5/2020	B2 ST131	ST131	fimH30	blaEC-5	blaTEM-1		dfrA17	aadA5				tet(A)	sul1	mph(A)				
8	c20	CARD	24-Sep	SEQ 5/2020	B2 ST131	ST131	fimH30	blaEC-5	blaCTX-M-27		dfrA17	aadA5	aph(3'')-Ib	aph(6)-Id		tet(A)	sul2					
9	c2-27	CARD	24-Sep	SEQ 5/2020	B2 ST131	ST131	fimH30	blaEC-5			dfrA17	aadA5										
10	L2-28	Longreach	26-Sep	SEQ 4/2020	B2 ST131	ST131	fimH30	blaEC-5	blaTEM-1		dfrA12	aadA2			aac(3)-IId		sul1	mph(A)				
11	L3-58	Longreach	26-Sep	SEQ 1/2021	B2	ST131	fimH30	blaEC-5	blaCTX-M-27		dfrA17	aadA5	aph(3'')-Ib	aph(6)-Id		tet(A)	sul1, 2	mphA				
12	M3a	Marlow	26-Sep	SEQ 4/2020	B2 ST131	ST131	fimH41	blaEC-5	blaCTX-M-15		dfrA17	aadA5					sul1	mph(A)	erm(B)			
13	M2-38	Marlow	26-Sep	SEQ 1/2021	B2 ST131	ST131	fimH30	blaEC-5														
14	p115	PONT	24-Sep	SEQ 4/2020	B2	ST131	fimH30	blaEC-5	blaTEM-215													
15	p62	PONT	24-Sep	SEQ 4/2020	B2 ST131	ST131	fimH30	blaEC-5	blaCTX-M-27													
16	p68	PONT	24-Sep	SEQ 4/2020	B2 ST131	ST131	fimH41	blaEC-5														
17	p71	PONT	24-Sep	SEQ 4/2020	B2 ST131	ST131	fimH41	blaEC-5	blaTEM-1													
18	p80	PONT	24-Sep	SEQ 4/2020	B2 ST131	ST131	fimH41	blaEC-5														
19	p88	PONT	24-Sep	SEQ 4/2020	B2 ST131	ST131	fimH22	blaEC-5														
20	R2-16	Reading	26-Sep	SEQ 4/2020	B2	ST131	fimH0	blaEC-5														
21	R2-27	Reading	26-Sep	SEQ 4/2020	B2	ST131	fimH41	blaEC-5	blaCTX-M-15	blaTEM-1												
22	R28	Reading	26-Sep	SEQ 4/2020	B2 ST131	ST131	fimH41	blaEC-5	blaTEM-1						aac(3)-IId							
23	R3-79	Reading	26-Sep	SEQ 4/2020	B2 ST131	ST131	fimH30	blaEC-5	blaCTX-M-15	blaOXA-1				aac(6')-Ib-cr	aac(3)-IIa	tet(A)			catB3			
24	p75	PONT	24-Sep	SEQ 4/2020	B2 ST131	ST131cp	fimH30	blaEC-5														
25																						
26	b32	Bris	19-Sep	SEQ 12/2020	D ST69	ST69	fimH27	blaEC-8									sul2					
27	b84	Bris	19-Sep	SEQ 12/2020	D ST69	ST69	fimH27	blaEC-8														
28	b103	Bris	19-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA12	aadA1	aph(3'')-Ia		aac(3)-IId	tetA	sul1,2,3	mdfA	cmlA1	floR	cmlA1	inuF
29	c39	CARD	24-Sep	SEQ 1/2021	B2	ST69	fimH27															
30	c3-16	CARD	24-Sep	SEQ 1/2021	D	ST69	fimH27	blaEC-8	blaTEM-1													
31	c3-30	CARD	24-Sep	SEQ 1/2021	D	ST69	fimH27	blaEC-8	blaTEM-215			aadA1	aph(3'')-Ia				sul3		metB	qnrs1		
32	c3-44	CARD	24-Sep	SEQ 1/2021	D	ST69	fimH483	blaEC-8	blaTEM-1C		dfrA12	aadA1	aph(3'')-Ib	aph(6)-Id		tet(A)	sul1	mphA	mdfA			
33	L116	Longreach	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA14					tet(A)		mph(A)				
34	L2-23	Longreach	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA5		aph(3'')-Ib	aph(6)-Id		tet(A)	sul2					
35	L2-26	Longreach	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8														
36	L2-31	Longreach	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1								sul1					
37	M33a	Marlow	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1													
38	M2-13	Marlow	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA17	aadA5	aac(3)-IId	aph(6)-Id		tet(A)	sul1, 2	mphA				
39	M2-15	Marlow	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8			dfrA17	aadA5	aac(3)-IId	aph(6)-Id								
40	M2-70	Marlow	26-Sep	SEQ 1/2021	D ST69	ST69	fimH47	blaEC-8														
41	p26	PONT	24-Sep	SEQ 1/2021	D	ST69	fimH27	blaEC-8														
42	p43	PONT	24-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA17	aadA5				tet(B)	sul1					
43	p63	PONT	24-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA17X3	aadA5		catA1		tetB	sul1					
44	p84	PONT	24-Sep	SEQ 1/2021	D	ST69	fimH27	blaEC-8	blaTEM-1				aph(3'')-Ib	aph(6)-Id		tet(34)	sul2					
45	p9	PONT	24-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1				aph(3'')-Ib	mphA		tet(A)	sul2					
46	R20	Reading	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA7		aac(3)-IId	aph(6)-Id		tet(A)	sul1, 2					
47	R4-41	Reading	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8	blaTEM-1		dfrA17	aadA5						catA1				
48	R4-64	Reading	26-Sep	SEQ 1/2021	D ST69	ST69	fimH27	blaEC-8														
49	b55	Bris	19-Sep	SEQ 12/2020	D ST69	ST69cmplx	fimH0	blaEC-8			dfrA1			aadA1			sat2					
50																						

Figure 5.13 Resistance gene complement of the prevalent phylotype B2 and D *E. coli* carriage isolate

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
51	b40	Bris	19-Sep	SEQ 4/2020	B2 ST69	ST127	fimH341	blaEC-5														
52	b76	Bris	19-Sep	SEQ 5/2020	B2	ST127	fimH2	blaEC-5								tet(B)						
53	b87	Bris	19-Sep	SEQ 5/2020	B2 ST131	ST127	fimH54	blaEC-5	blaOXA-1				aadA1			tet(B)						
54	b2-12	Bris	19-Sep	SEQ 5/2020	B2 ST95	ST127	fimH100	blaEC-5														
55	c29	CARD	24-Sep	SEQ 5/2020	B2	ST127	fimH126	blaEC-5														tet(34)
56	c2-58	CARD	24-Sep	SEQ 5/2020	B2	ST127	fimH2	blaEC-5														
57	c3-32	CARD	24-Sep	SEQ 1/2021	B2	ST127	fimH2	blaEC-5	blaTEM-1				aph(3'')-Ib	aph(6)-Id				mdfA		catA	tet(B)	
58	c3-45	CARD	24-Sep	SEQ 1/2021	B2	ST127	fimH54	blaEC-5														tet(B)
59	L31	Longreach	26-Sep	SEQ 5/2020	B2	ST127	fimH2	blaEC-5	blaTEM-150													
60	L115	Longreach	26-Sep	SEQ 4/2020	B2	ST127	fimH54	blaEC-5	blaOXA-1				aadA1			tet(B)				catA1		
61	M26b	Marlow	26-Sep	SEQ 1/2021	B2	ST127	fimH2	blaEC-5														
62	M2-26	Marlow	26-Sep	SEQ 4/2020	B2	ST127	fimH2	blaEC-5														
63	p20	PONT	24-Sep	SEQ 4/2020	B2	ST127	fimH121	blaEC-5														
64	p33	PONT	24-Sep	SEQ 10/202(	B2	ST127	fimH181	blaEC-5														
65	p36	PONT	24-Sep	SEQ 10/202(	B2	ST127	fimH462	blaEC-5	blaTEM-150													
66	p37	PONT	24-Sep	SEQ 4/2020	B2	ST127	fimH181	blaEC-5														
67	p75	PONT	24-Sep	SEQ 4/2020	B2	ST127	fimH41	blaEC-5	blaTEM-1		aph(6)-Id		aph(3'')-Ib			tet(A)	sul2					
68	p90	PONT	24-Sep	SEQ 1/2021	B2	ST127	fimH136															
69																						
70	b61	Bris	19-Sep	SEQ 4/2020	B2 ST73	ST73	fimH27	blaEC-5														
71	b95	Bris	19-Sep	SEQ 5/2020	B2 ST73	ST73	fimH12	blaEC-5														
72	c23	CARD	24-Sep	SEQ 5/2020	B2 ST73	ST73	fimH30	blaEC-5														
73	c2-51	CARD	24-Sep	SEQ 1/2021	B2	ST73	fimH30	blaEC-5					dfrA17	aadA5			sul1		mph(A)			
74	c2-60	CARD	24-Sep	SEQ 1/2021	B2	ST73	fimH103	blaEC-5												tet(A)		
75	c3-11	CARD	24-Sep	SEQ 1/2021	B2	ST73	fimH10	blaEC-5														
76	p1	PONT	24-Sep	SEQ 4/2020	B2 ST73	ST73	fimH10	blaEC-5														
77	p103	PONT	24-Sep	SEQ 4/2020	B2 ST73	ST73	fimH103	blaEC-5	blaTEM-150													
78	p34	PONT	24-Sep	SEQ 4/2020	B2 ST73	ST73	fimH10	blaEC-5	blaTEM-1								sul2					
79	R3-64	Reading	26-Sep	SEQ 5/2020	B2 ST73	ST73	fimH504	blaEC-5	blaTEM-1													
80	R3-65	Reading	26-Sep	SEQ 1/2021	B2 ST73	ST73	fimH10	blaEC-5	blaTEM-1													
81	R3-66	Reading	26-Sep	SEQ 5/2020	B2 ST73	ST73	fimH10	blaEC-5	blaTEM-1													
82	R4-19	Reading	26-Sep	SEQ 1/2021	B2 ST73	ST73	fimH0	blaEC-5														
83	R4-36	Reading	26-Sep	SEQ 1/2021	B2 ST73	ST73	fimH12	blaEC-5	blaSHV-1			aadA1					sul1					
84																						
85	b30	Bris	19-Sep	SEQ 1/2021	B2	ST95	fimH27	blaEC-5								tet34						
86	b77	Bris	19-Sep	SEQ 4/2020	B2 ST95	ST95	fimH41	blaEC-5	blaTEM-235				aph(6)-Id	aph(3'')-Ib		tet(34)	sul2					
87	b78	Bris	19-Sep	SEQ 4/2020	B2 ST95	ST95	fimH27	blaEC-5	blaTEM-150							tet(34)						
88	b92	Bris	19-Sep	SEQ 5/2020	B2 ST95	ST95	fimH30	blaEC-5								tet(34)						
89	c27	CARD	24-Sep	SEQ 5/2020	B2 ST95	ST95	fimH27	blaEC-5														
90	c3-25	CARD	24-Sep	SEQ 1/2021	B2	ST95	fimH30	blaEC-5								tet(34)						
91	c3-27	CARD	24-Sep	SEQ 1/2021	B2	ST95	fimH41	blaEC-5								tet(34)						
92	c3-43	CARD	24-Sep	SEQ 1/2021	B2	ST95	fimH27	blaEC-5	blaTEM-1		dfrA12	aadA1,2				tet(B, 34)	sul3	cmlA1	mefB			
93	M2-4	Marlow	26-Sep	SEQ 1/2021	B2 ST95	ST95	fimH41	blaEC-5								tet(34)						
94	p105	PONT	24-Sep	SEQ 4/2020	B2 ST69	ST95	fimH41	blaEC-5	blaTEM-150							tet(34)						
95	p48	PONT	24-Sep	SEQ 4/2020	B2 ST95	ST95	fimH41	blaEC-5	blaTEM-1							tet(34)	sul1					
96																						
97	B3	Bris	19-Sep	SEQ 4/2020	B2	ST404	fimH27	blaEC-5	aadA1		dfrA1		sat2_gen									
98	c2-26	CARD	24-Sep	SEQ 1/2021	B2	ST404	fimH27	blaEC-5	blaTEM-1								sul2					
99	L3-80	Longreach	26-Sep	SEQ 1/2021	B2	ST404	fimH27	blaEC-5			dfrA1		aadA1	sat2								
100	p15	PONT	24-Sep	SEQ 4/2020	B2	ST404	fimH27	blaEC-5														
101	p67	PONT	24-Sep	SEQ 4/2020	B2	ST404	fimH27	blaEC-5	blaTEM-1								sul2					
102	p77	PONT	24-Sep	SEQ 4/2020	B2	ST404	fimH27	blaEC-5	blaTEM-1								sul2					
103																						

Figure 5.14 Resistance gene complement of the prevalent phylotype B2 and D *E. coli* carriage isolate

Other aminoglycoside-modifying genes were also common in these MLST as well as genes conferring macrolide resistance. Interestingly the *tet34* allele was often found in ST95 isolates, this has important implications for ST95 isolates that may carry carbapenemase genes as *tet34* is known to confer tigecycline resistance, one of the few options for carbapenem-resistant isolates. Other MLST types harboured far fewer resistance mechanisms (**Figures 5.13 and 5.14**). Because most of the other phylotypes (other than B2 and D) were not sequenced we cannot comment on the genetic basis of resistance in those isolates. Our study did not find any carbapenem resistance encoding genes.

5.4 Discussion

Data collected by the public health authorities in all nations within the UK has indicated that bacterial sepsis has been increasing year on year since voluntary records were initiated in the 1990's (Abernethy et al. 2017). This appears to be a common thread throughout Europe, and Northern America and has severe implications in outcomes from secondary infections such as those caused by a prior viral infection such as the current COVID-19 epidemic. Heading the list are bacteraemia caused by common human-associated organisms such as *Escherichia coli* and *Staphylococcus aureus*. *E. coli* itself causes more sepsis cases than any other organism in the UK as well as more urinary tract infections than any other organism (Daoud and Afif 2011). This is partly due to its near universal carriage by all members of our population as opposed to other organisms such as *S. aureus* and *K. pneumoniae* (both typically c. 20% carriage) (Livermore et al. 2007). Mandatory surveillance of *E. coli* bacteraemia began in the UK in 2011 with numbers indicating a doubling of the sepsis rate since that time. The underlying reasons behind this increase are enigmatic and cannot simply be due to an ageing population since the

rate of infection has been increasing in all age groups. As mentioned before, *E. coli* is a hugely diverse species with only 1,000 of the typical individual complement of 5,000 genes shared by all members of the species (Land et al. 2015). This indicates that *E. coli* easily loses and gains genetic information. Such genetic fluidity has led to a major division within the species between individual members that are either commensal rarely causing disease (typically only in immunocompromised individuals) and pathogens that are commonly associated with severe infection. Studies on the numbers of *E. coli* strains carried by individuals indicate that most people tested carry an individual strain (60%), followed by approximately 30% of people who carry 2 strains and a minority of individuals who carry multiple strains. Thus, the carriage of either a commensal or pathogenic type can lead to a dichotomy of outcome in terms of propensity to UTI and bacteraemia. On a population level, the proportion of individuals carrying pathogenic *E. coli* strains could easily have a dramatic effect on national *E. coli* sepsis rates. We thus hypothesised that the consistently higher *E. coli* sepsis rates between Wales and England could be due to a higher prevalence and carriage rate of pathogenic types between these two nations. To test this hypothesis, we sampled human sewage on arrival at 6 waste-water sewage stations along the M4 motorway corridor that links England to Wales. The sites were chosen close to either end of the corridor. As a comparator, we also collected sewage, in the same week, from Bangladesh since we have consistently found low rates of pathogenic ExPEC *E. coli* types in South Asia. We found a statistical difference between the prevalence of the main pathogenic B2 phylotype but not for phylotype D, between the sites chosen at East and West ends of the M4, which was also statistically relevant for comparison between sites within Wales and within England. The difference between the prevalence of pathogenic organisms in sewage clearly reflected the overall difference between sepsis rates in these two

Nations. Encouraged by this result we further Whole genome sequenced all *E. coli* isolates belonging to pathogenic phylotypes B2 and D. The deduction of the MLST types of these organisms revealed the surprising overall result that the most prevalent *E. coli* ST at each site belonged to only a handful of ST's belonging to the Phylotype B2 (ST131, ST73, ST127, ST95, ST404) as well as ST69 belonging to phylotype D. At each site these MLST types represented an average prevalence of 60% (range 51-70%) of all isolates belonging to phylotype B2. This information dramatically reflects the known number and proportion as well as the identity of *E. coli* MLST types isolated from *E. coli* bacteraemias in our hospitals. The rational explanation of these results is that *E. coli* sepsis is dramatically influenced by the carriage rates of pathogenic *E. coli* in the community and suggests that a reasonable explanation for the rising *E. coli* sepsis rates in the western world is the rising carriage rates of virulent *E. coli* types in the community. Analysis of *E. coli* types from Bangladesh indicated an absence of pathogenic B2 types, an observation that we have made previously for several S. Asian nations including Pakistan, Bangladesh, and India. In fact, we and others have found that different nations have different prevalent *E. coli* MLST types (see references in the introduction 5.1). This interesting observation may underlie the particularly low sepsis rates found within the London area which are the lowest found in the UK and may be related to the fact that 35% of people living in London are recent immigrants who were not born in the UK (Sturge 2023). It is also an interesting fact that though S. Asia is plagued with shockingly high rates of antibiotic resistance in *E. coli*, this tends to generally be in non-pathogenic types of *E. coli* that rarely cause ExPEC disease. This additionally is mirrored in the recent finding that the carriage rate of ESBL *E. coli* is only 8% in Wales as compared to 17% in London (Day et al. 2019). The differences in *E. coli* types carried in different nations must be a natural phenomenon and we have

preliminary information that this is related to both sanitation and bacteriophage predation. Since it is likely that differences in *E. coli* carriage between different nations are due to natural phenomena this suggests that these natural processes can be manipulated, and we would suggest that this should be a priority area of research if the UK government is serious about reducing *E. coli* bacteraemia rates for the UK. Similarly, the differences probably extend to other pathogens and likely are important to the understanding of the widely varying death rates due to secondary bacterial infection in different nations as has been highlighted by the COVID-19 pandemic.

Our results indicate that sepsis rates are hugely influenced by carriage of only a few *E. coli* ST. At the top of this list for the UK is ST131 and our information indicates that several varieties of ST131 are commonly carried in the UK, including of various clades that have been differentiated previously by possession of different *fimH* alleles. We found all these clades in our study and as shown previously we found that the $bla_{\text{CTX-M-15}}$ gene is most often associated with the *fimH30* allele. Interestingly we found that $bla_{\text{CTX-M-27}}$ was as common in our collection as $bla_{\text{CTX-M-15}}$ which may suggest a shift in dominant ESBL since the prior study of Day *et al* 2019. Importantly 5 years have passed since these two collections ie 2013 for Day *et al* and late 2019 for this current study. Importantly we found in this study that antibiotic resistance is only closely associated with ST131 and ST69 and not so much with other pathogenic types. This has been observed before and highlights the importance of studies such as this one which selected isolates without any antibiotic bias. In fact, the common carriage of other pathogenic *E. coli* MLST would not be discovered without this type of study. This is highlighted by the national importance of ST127. The discovery of the very high prevalence of ST127 in our collection especially at the westerly locations indicates that this MLST is very important to UK sepsis and has been

increasing in prevalence in recent years. This MLST has only very recently been identified in the UK (Gibreel et al. 2012) and its importance is dramatically highlighted by this study but was rarely found in the Day *et al* study because of the antibiotic bias of that study.

Remarkably, Theodore Escherich once said **"It would appear to be a pointless and doubtful exercise to examine and disentangle the apparently randomly appearing bacteria in normal faeces and the intestinal tract, a situation that seems controlled by a thousand coincidences".** We believe we have demonstrated the value of examining and disentangling these differences, if only for the organism which is his namesake, in the hope of helping to combat rising sepsis rates.

Chapter 6 Comparing the Virulence of *E. coli* ST131 Based on the Genetic Location of *bla*$_{\text{CTX-M-15}}$ in the *Galleria mellonella* Model

6.1 Introduction:

6.1.1 Prior studies using *G. mellonella* to study *E. coli.*

The mouse model is one of the research tools that is used most frequently while investigating microbial illnesses. Also, studies with human volunteers that would best imitate bacterial virulence are presently not possible due to ethical concerns (Nataro et al. 1995). Using rodents as infection models does, however, present challenges on several levels including ethical, financial, and logistical. Greater wax moths, also known as honeycomb moths, of the species *Galleria mellonella* have been utilised as an alternate model for research on microbial diseases. They belong to the Lepidoptera and Pyralidae families (snout moths) (Common 1975). In actuality, the caterpillar larvae or wax worm are employed as an animal model, not the adult moth. *G. mellonella* larvae are less expensive and simpler to keep than typical mammalian model hosts since they do not require specialised laboratory equipment (Ramarao et al. 2012). In addition, *G. mellonella* usage does not require ethical approval, and their short lifespan makes them perfect for investigations (Tsai et al. 2016a). This has served as an efficient model for researching the pathogenesis, pathogenicity, and antibiotic treatment of a wide variety of bacterial and fungal infections (Williamson et al. 2014a). The model has a number of advantages, including the capability to incubate larvae over a range of temperatures (from 15 to 37 degrees Celsius), adaptability to both low- and high-throughput experiments, relative ease of maintenance and handling, and the fact that the bacteria/drug inoculum can be accurately quantified prior to infection (Ciesielczuk et al. 2015a).

Insects may die from human infections, and their mutations can be used to assess their pathogenicity (Taszłow et al. 2017). Knockout, knockdown, or overexpression of certain genes is carried out in the investigation for virulence factors, and mutant

strains of these genes are then tested for their pathogenicity against *G. mellonella* (Harding et al. 2013; Giannouli et al. 2014). Notably, this is conceivable not only because the immune systems of mammals and insects are comparable, but also because the immune systems of mammals and insects function similarly (Desalermos et al. 2012).

The pathogen of interest is introduced into the chosen larvae through the process of inoculation. One of the three available methods of infection can be used to accomplish the process of inoculation.

1. Topical application. Scully and Bidochka (Scully and Bidochka 2005) provide an explanation of this methodology. The distribution of the inoculum using the topical application, in which the inoculum is deposited on the cuticle of the larvae, requires relatively little effort on the part of the user. By breaking through the cuticle of the insect, the microbe can gain access to the larvae. This delivery method has the disadvantage that it prevents knowing the number of infecting fungal cells.
2. Oral delivery. One way that *G. mellonella* becomes infected is by inoculation through feeding (Fedhila et al. 2010). In this method, the pathogen is mixed with pollen in an equal amount, and the mixture is then put on a plate with *G. mellonella* larvae. The infection can get into the larvae and spread through the population when they eat it. The ingestion method gives a form of infection that is similar to how fungi normally get into insects. As a result, the full range of the natural hosts' defences are activated.
3. Injection. Infection of *G. mellonella* occurs when the inoculum is injected into the hemocoel of the organism. At the very end of the left pro-leg is where the inoculation injection is given. If more injections are required, several injections may be given. However, a fresh injection site must be utilised for each

successive injection (Fuchs et al. 2010). The injection method has the advantages of knowing the exact quantity of the pathogen injected to the hemocoel into initiate the infection process.

Intestinal chitin provides *G. mellonella* with a layer of defence against the environment. The cuticle also protects the internal organs that originated from the ectoderm, such as the trachea, foregut, and hindgut, preventing pathogens from entering the body (Vertyporokh and Wojda 2017). When these barriers are breached, the systems of defence are set into motion. The activation of signalling pathways that regulate the generation of antimicrobial peptides is one of them. These peptides, when released into the haemolymph, act against the bacteria that are causing the infection (Wojda et al. 2020). Following damage or infection, the phenol oxidase system is released from within the haemocytes (oenocytoides). These have prophenol oxidase, which is a pro-enzyme that must be broken down by serine proteases before it can be converted into the active enzyme (PO). Parallel to this, serpins, which are inhibitors of serine proteases and are a component of the phenoloxidase complex, stop enzyme hyperactivation, which can be harmful to the host owing to the generation of free radicals during the production of melanin (Bidla et al. 2009; Demir et al. 2012). The production of the dark melanin pigment from tyrosine under the catalytic influence of prophenol oxidase is essential to the success of this process. Melanin can attach itself to the surface of microbes, making the process of removing them easier. Melanisation frequently occurs in conjunction with the process of haemolymph coagulation. This process makes the clot more rigid and inhibits haemolymph efflux until the epidermis is regenerated (Li et al. 2002).

Table 6.1. Components of the *G. mellonella* innate immune (Tsai et al. 2016a)

a) cellular response	
haemocyte	Prohemocytes, plasmatocytes, granular cells, coagulocytes, spherulocytes, oenocytoids
b) *humoral response*	
Opsonins	apolipophorin-III (apoL-III), peptidoglycan recognition proteins (PGRPs), cationic protein 8 (GmCP8), hemolin
antimicrobial peptides (AMPs)	Lysozyme, cecropin, moricin-like peptides, gloverin, galiomycin, gallerimycin, Galleria defensin, Gm proline-rich peptides 1 and 2, Gm anionic peptide 1 and 2, inducible serine protease, inhibitor 2, heliocin-like peptide, x-tox, Gm apolipophoricin
Melanisation	phenoloxidase

More recently, a health index scoring system (**Table 6.2**) was created. This method evaluates the condition of the larvae's health by giving scores based on four primary observations: the larvae's movement, the construction of cocoons, the development of melanin, and their ability to survive (Loh et al. 2013). Melanisation of the larvae often begins with the appearance of prominent black patches on larvae that are cream coloured. The completion of melanisation, which results in black larvae, is followed shortly thereafter by the larvae's demise (Tsai et al. 2016a).

Table 6.2. The *G. mellonella* health Index Scoring System (Loh et al. 2013)

Category	Description	Score
Activity	no movement	0
	minimal movement on stimulation	1
	move when stimulated	2
	move without stimulation	3
Cocoon formation	no cocoon	0
	partial cocoon	0.5
	full cocoon	1
Melanisation	Black larvae	0
	≥3 spots on beige larvae	1
	<3 spots on beige larvae	2
	No melanisation	3
Survival	dead	0
	alive	2

In addition to determining the crude mortality rate, it is simple to monitor both the innate immune responses and the cellular damage that occurs after an infection (Lee et al. 2004). It is simple to measure the generation of melanin, which *G. mellonella* induces during the acute phase response and is comparable to the human complement cascade. In the meantime, the cellular damage marker lactate dehydrogenase (LDH), which is generated by cells after apoptosis, may be used to identify damaged cells (Leuko and Raivio 2012; Williamson et al. 2014b). Thus, the generation of LDH and melanin may both be utilised as quantitative measures of immunogenicity and responsiveness.

The *G. mellonella* infection model has been the subject of more than a thousand publications on PubMed **(Table 6.3**). Of these, more than 900 were published in only the last five years 2017–2022, which is evidence of the growing popularity of this infection model (Tsai et al. 2016b). In 2012, research was published that was the first of its kind to describe using an infection model based on *G. mellonella* to investigate pathogenic *E. coli* (Ciesielczuk et al. 2015b). Leuko and Raivio found that enteropathogenic *E. coli* (EPEC) could kill *G. mellonella* larvae in a dose-dependent

manner, with a lethal dose of 50 (LD50) value of 2.57×10^3 CFU at 48 hours after infection. This value was determined using a standard curve. After being injected into the hemocoel, the bacteria vanished not long afterward and became confined to melanised capsules. One study using the *G. mellonella* model has shown a notable link between the virulence gene repertoire of extraintestinal pathogenic *E. coli* (ExPEC) and the virulence capability of the bacteria. ExPEC strains with a greater number of virulence genes led to a considerably more rapid death of the larvae (Clermont et al. 2009). In another study that was quite comparable to this one, forty well-characterised ExPEC strains were investigated using a *Galleria mellonella* infection model. They measured larval survival, melanisation, and cell damage, and noted increased virulence in isolates from community-associated infections, complicated UTIs, and urinary-sourced bacteraemia, particularly in isolates belonging to the ST131 lineage (Ciesielczuk et al. 2015b).

The presence of *E. coli* bacteraemia has been shown to be associated with an increased risk of death in humans ranging from 17 to 61 percent (Melzer and Petersen 2007; Schlackow et al. 2012). UTIs have been identified as the most common cause, and the incidence of the disease is growing at an alarming rate each year in the United Kingdom (www.hpa.org.uk). One third of women will have experienced at least one UTI before the age of 24, and around fifty percent of women will be treated for a symptomatic UTI at some point throughout their career (Excellence. 2021). In males who are young or middle-aged and otherwise healthy, UTI is an extremely rare condition. Those who are older males and those who have urinary catheters that are permanently implanted have a higher incidence rate (Excellence. 2018). This trend toward increased virulence among *E. coli* urosepsis strains is reflected in *G. mellonella*, whereby, larvae mortality was found to be significantly higher in strains associated with urosepsis compared to those from

patients with an uncomplicated UTI or from a non-urinary source of bacteraemia. This was found to be the case when comparing strains from patients with an uncomplicated UTI or from a non-urinary source of bacteraemia. In addition, the fact that ST131 isolates were responsible for the highest fatality rate indicates that these particular strains are likely substantially more virulent than the other major lineages. Infections in the urinary tract and bloodstream are common places where ST131 is found (Ciesielczuk et al. 2015b).

Although the *G. mellonella* infection model has been used to study the virulence of a variety of clinically significant bacterial and fungal diseases, it was only recently recognised as a feasible model for studying ExPEC virulence (Williamson et al. 2014a). Williamson *et al*. (2014) conducted an analysis of nine strains recovered from bloodstream infections and discovered a connection between a high aggregate virulence score and larval death (Williamson et al. 2014a).

Table 6.3. Publications that using *Galleria mellonella* as a model to study *E. coli*

References	Year of publication	Type of microbes	Number of isolates	Comments
Leuko, S. *et al*.(Leuko and Raivio 2012)	2012	Enteropathogenic *E. coli* (EPEC).	7	The study shows that mutations that activate the Cpx pathway have a substantial influence on the bacteria's capacity to produce a deadly infection, and this is connected with an inability to grow in vivo. The fact that the bacterium was unable to grow in vivo served as proof of this. The findings of this study suggest that inducing the Cpx envelope stress response can result in virulence issues.
Alghoribi, M. F, *et al*.(Alghoribi et al. 2014)	2014	*E. coli* (sequence types ST69, ST127, ST73, ST131).	71	This study investigated 71 isolates from five leading strains of UPEC. ST131 was the least virulent of the other groups. However, ST69 and ST127 showed higher mortality rates.
Williamson, *et al*.(Williamson et al. 2014a)	2014	Extraintestinal pathogenic *E. coli* (ExPEC).	11	This study investigates molecularly characterised ExPEC strains to assess their virulence and the feasibility of using *G. mellonella* as a model. The study concluded a correlation between the number of virulence genes and larval surviving.
Ciesielczuk *et al*.(Ciesielczuk et al. 2015b)	2015	Extraintestinal pathogenic *E. coli* (ExPEC).	40	This study investigated 40 ExPEC strains by measuring mortality, immune response, and cell damage. It concluded that the mortality rate was similar between urinary and bloodstream isolates. However, it was higher for community-associated infections complicated by UTIs and urinary-sourced bacteraemia. Interestingly, isolates of ST131 and those possessing *afa/dra, ompT*, and serogroup O6 were also associated with higher virulence.
Torres, M. P. *et al*.(Torres et al. 2016)	2016	*E. coli* and *K. pneumoniae.*		The aim of this study is to determine whether immunosuppression with dexamethasone-phosphate might be utilised in an in vivo infection model utilising *G. mellonella.*
Clermont *et al*.(Clermont et al. 2009)	2016	*E. coli.*	486 of commen-sal and bacterae-mic strains of *E. coli.*	This study compares phylogenetic grouping, O-type, virulence factor, and antibiotic resistance. It concluded that two levels of specialisation from commensal strains were demonstrated in the bacteraemic strains: resistance to antibiotics in all cases and virulence for those of urinary tract origin.
Heitmueller, M *et*	2017	Uropathogenic *E. coli* (UPEC).		This study compares UPEC and ABU strains using *G. mellonella* to determine which

References	Year of publication	Type of microbes	Number of isolates	Comments
al.(Heitmueller et al. 2017)				epigenetic mechanism can explain the difference between pathogen and commensal *E. coli.*
Scalfaro, C. *et al*.(Scalfaro et al. 2017)	2017	*Salmonella enterica*, EPEC sero type O127:H6, and *L.monocytogenes* serotypes 4b	3	*G. mellonella* larvae were used in this study to see if it was possible to assess probiotic bacteria's defence against gastrointestinal bacterial infections.
Guerrieri, et al.(Ménard et al. 2021)	2019	Enteroaggregative *E. coli* (EAEC).	20	This study compared typical and atypical based on the presence of the *AggR* regulon. They conclude that both groups are virulent in the *G. Mellonella* model.
Vergis, J. *et al*.(Vergis et al. 2019)	2019	Enteroaggregative *E. coli* (EAEC).		This study shows the efficacy of Indolicidin against resistant *E. coli* using *G. mellonella*. It concludes that indolicidin is an effective antimicrobial candidate against MDR-EAEC.
Wang, L. *et al*.(Wang et al. 2020)	2020	*E. coli.*	4	This study provides information on the use of bacteriophages and antibiotics in the treatment of resistant *E. coli*.
Petronio, P. *et al*.(Petronio et al. 2020)	2020	Uropathogenic *E. coli.*	5	This study tested berberine in vivo and in vitro to assess the activity of berberine on uropathogenic *E. coli* strains. It concludes that the use of berberine increases the larval survival of infected UPEC.
Mukherjee, K. *et al*.(Mukherjee et al. 2020)	2020	Uropathogenic and commensal *E. coli.*		This study compares UPEC and ABU strains using *G. mellonella.* It concluded that microRNAs are involved in the epigenetic reprogramming of the innate immune response in *G. mellonella* larvae. This allows the larvae to differentiate between commensal and pathogenic strains of *E. coli*.
Antoine, C. *et al*.(Antoine et al. 2021)	2021	Avian pathogenic *E. coli.*	11	This study investigates the treatment using bacteriophages against *E. coli* K1. It concluded that treatment with phages improves the larvae's chances of surviving.
Virgis, J. *et al*. (Vergis et al. 2021)	2021	Enteroaggregative *E. coli* (EAEC).		This study evaluated the antibacterial effectiveness of Cecropin A (1-7)-Melittin (CAMA) against three MDR *E. coli* (MDR-EAEC).
Ram, V. P. *et al*.(Prasastha Ram et al. 2022)	2022	Enteroaggregative *E. coli* (EAEC).		This study looks at how well nanosilver-entrapped cinnamaldehyde (AgC) kills multidrug-resistant strains of enteroaggregative E. coli (EAEC), how stable it is in the lab, and if it is safe.

References	Year of publication	Type of microbes	Number of isolates	Comments
Erol, H. B. *et al*.(Erol et al. 2022)	2022	ESBL *E. coli.*		This study investigated potential therapeutic phages against *E. coli*. In conclusion, phages were found to have potential for the treatment of *E. coli* infections.

6.1.2 E. coli growth rate

In microbiology, growth rates have been utilised for a very long time to measure phenotypic traits. In the study of experimental evolution, growth rates have frequently been employed as an indicator to measure (Hall 1978; Dykhuizen and Dean 1990). The development of high-throughput analysis for growth rate measurements has been made possible by the invention of automated microtiter plate readers. This development has led to a resurgence in the application of growth rate data in a wide variety of fields, such as the regulation of gene expression and numerous areas of microbiology, including the antibacterial activities of biological fluids (Fehér et al. 2012).

6.2 Results

The comparison of the growth rates of *E. coli* ST131 strains with their resistance genes solely on plasmidic locations as compared to both plasmid and chromosomal positions showed no significant difference in the growth rates of these two groups (each group consisted of six individual strains which were run in triplicates). Statistical analysis yielded a *P* value of 0.543, indicating no statistical differences. However, the comparison between growth rates of strains having a chromosomal only position as compared to the two groups containing plasmid locations indicated that strains with a sole chromosomal location grew slightly faster. This was statistically verified with a *P* value of <0.05 (**Figure 6.1**).

The survival of larvae was measured to see whether there was a link between the genetic location of the *bla*$_{CTX\_M\_15}$ gene and pathogenicity following the challenge of 17 *E. coli* strains belonging to one MLST type, ST131. In Chapter 3, S1-PFGE was performed on these isolates, indicating that in *E. coli* isolated from human sources, the transposition of the *bla*$_{CTX\text{-}M\text{-}15}$ gene into a chromosomal site was a common event. Forty-nine percent of the studied isolates had the *bla*$_{CTX\text{-}M\text{-}15}$ gene present on the chromosome in isolates collected from faecal sources and 57% from bacteraemia sources, with 42 percent of the isolates having the gene present on plasmids. This was observed to be present at a significantly higher frequency in isolates recovered from bacteraemia as opposed to carrier sources such as faeces and sewage. The transposition event had occurred in a wide variety of *E. coli* sequence types.

In the *G. mellonella* infection model, measuring microbial virulence normally takes place within three days, and the survival rate at various time periods is the end point that is used most of the time. It is possible to determine the half-maximal fatal dose, often known as the lethal dose (LD50), by inoculating larvae with a range of different doses. Two different doses were used on larvae 10^5 and 10^6 CFU/larvae to measure the *E. coli* ST131 virulence. The isolates were divided into three groups based on the location of the resistance gene *bla*$_{CTX\_M\_15}$.

The first group consisted of isolates of *E. coli* ST131 with the *bla*$_{CTX\text{-}M\text{-}15}$ gene found in two genetic locations, i.e., both plasmidic and chromosomal. For this group, when infected with 10^5 CFU *E. coli* ST131/larvae, the killing of *G. mellonella* ranged from 0% to 13% (mean 0.038) for individual isolates, compared to 0% to 37% (mean 0.088) when infected with 10^6 CFU *E. coli* ST131/larvae. The second group consisted of strains that only carried the *bla*$_{CTX\text{-}M\text{-}15}$ gene in a plasmidic location. When these strains were used to infect *G. mellonella* at 10^5 CFU *E. coli*

ST131/larvae, they showed killing of *G. mellonella* from 0% to 28% (mean 0.055), compared to 0% to 25% (mean 0.116) when infected with 10^6 CFU/larvae. The final group consisted of isolates that carried the *bla*$_{CTX\text{-}M\text{-}15}$ gene only in a chromosomal genetic location. When infected with 10^5 CFU *E. coli* ST131/larvae, they displayed killing rates of *G. mellonella* ranging from 0% to 26% (mean 0.106) and from 0% to 35% (mean 0.193) when infected with 10^6 CFU *E. coli* ST131/larvae. Overall, there appeared to be little difference in mortality rates between the different groups based on *E. coli* strains that possess *bla*$_{CTX\_M\_15}$ in different genetic locations.

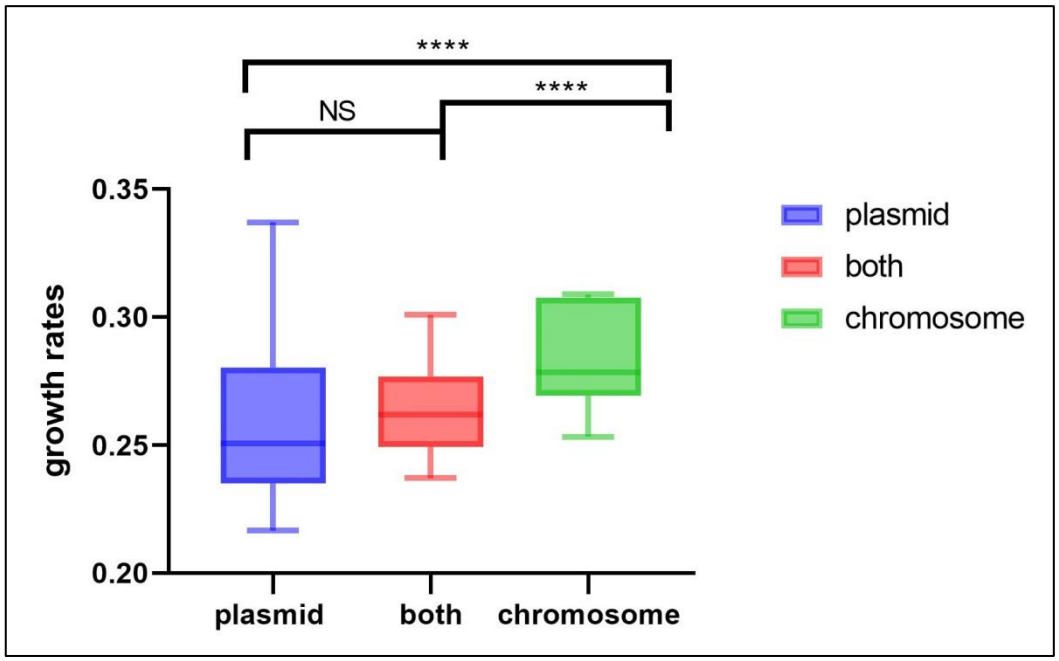


Figure 6.1. The growth rates of three different groups of *E. coli* based on the genetic location of *bla*$_{CTX\text{-}M\text{-}15}$ gene

NS stands for not significant P>0.05 and **** stands for P value less than 0.0001 (Prism 5.04).

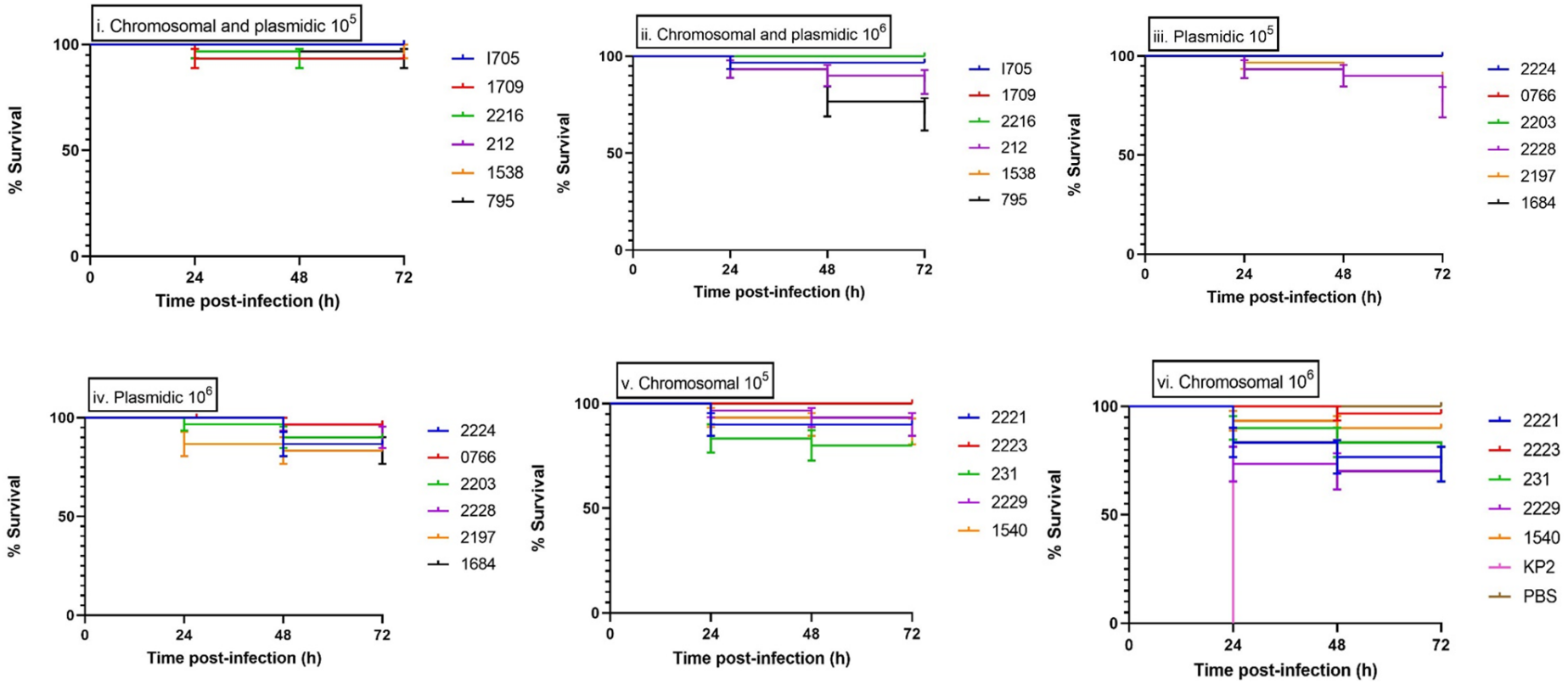


Figure 6.2. Larvae survival of *E. coli* ST131 strains with *bla*$_{CTX\_M\_15}$ gene in different genetic locations

Percentage survival was calculated from 3 separate killing assays 72-hours post-injection. Three groups with two different concentrations of doses i) 10^5 CFU/larvae of ST131 strains with their resistance genes *bla*$_{CTX\_M\_15}$ in both the chromosome and the plasmid. ii) 10^6 CFU/larvae of ST131 strains with their resistance genes *bla*$_{CTX\_M\_15}$ in both the chromosome and the plasmid iii) 10^5 CFU/larvae of ST131 strains with their resistance genes *bla*$_{CTX\_M\_15}$ in the plasmid iv) 10^6 of ST131 strains with their resistance genes *bla*$_{CTX\_M\_15}$ in both the chromosome and the plasmid v) 10^5 of ST131 strains with their resistance genes *bla*$_{CTX\_M\_15}$ in the chromosome vi) 10^6 CFU/larvae of ST131 strains with their resistance genes *bla*$_{CTX\_M\_15}$ in both the chromosome and the plasmid and positive control KP2 and negative control PBS.

6.3 Discussion

The chromosomal location appears to be important for pathogenicity. We observed in the study of the Department of Health in Chapter 3 that in this large group of isolates, 57.6% of chromosomally encoded resistance was found in bacteraemia isolates. This is compared to 44.1% of chromosomally encoded resistance found in faecal isolates and 31% of chromosomally encoded resistance found in sewage isolates. It is possible that this is because bacteraemic isolates grow more rapidly and are therefore better able to survive in the hostile environment of human blood. Also, PFGE analysis of the UK ESBL *E. coli* study found that the chromosomal carriage of the $bla_{CTX\text{-}M\text{-}15}$ resistance gene was surprisingly common. The strains with their resistance genes found on the chromosome grew faster in the growth experiment than those belonging to the other two groups. This is probably because movement to the chromosome is often associated (90%) with the loss of the donating plasmid. The loss of a large plasmid would likely be helpful for the bacterial cell since the bacterium effectively has lost a parasite that drains the bacterium of resources, and this advantage manifests itself in an overall shorter replication time, i.e., a faster growth rate.

Based on the observation that antibiotic resistant *E. coli* isolates recovered from bacteraemia more often had chromosomally encoded resistance genes than faecal or sewage derived isolates, we envisioned that a chromosomal location was likely a virulence factor. The successful growth experiments gave us an understanding of how this virulence factor might operate, i.e., an increased growth rate and therefore perhaps an ability to survive more easily in the blood. We therefore designed our *Galleria* experiment in this chapter to explore *G. mellonella* larval survival after infection with seventeen *E. coli* ST131 isolates harbouring the $bla_{CTX\text{-}M\text{-}15}$ in different

genomic locations (**Table 2.7**). This study was aided by the fact that all ST131 isolates are essentially clonal, and therefore the background genetics of each isolate are very similar, with the main differences being the number of plasmids carried and the position of the antibiotic resistance genes. The results revealed little difference between the three groups with slightly more killing at the 10^6 chromosomal group. However, all tested *E. coli* strains were generally of low virulence, as determined from our experiment. All strains belong to one sequence type, which is ST131, which may explain the little differences in the results because it is one clone other than plasmid contents and chromosomal locations. In a study by Alghoribi *et al*. investigating 71 isolates to show the genetic differences of five leading lineage of UPEC *E. coli* (e.g., sequence types 69, ST73, ST95, ST127, and ST131), it concluded that ST69 and ST127 were highly virulent among other strains and ST131 had generally lower virulence (Alghoribi et al. 2014). In fact, in a study of antibiotic resistant *E. coli* distributed throughout the UK, we found that ST131 caused more bacteraemia events than any other strain. However, the number of events could perhaps be better explained by the fact that this strain was carried by more people in the UK than any other strain and was therefore more often able to cause disease than its innate ability to do so.

One hypothesis can be offered to explain the little difference in lethality between the strains from the different groups: virulence has been shown to be related to the gain/loss of genes in ExPEC strains. In a study conducted by Ciesielczuk *et al* (Ciesielczuk et al. 2015b) comparing the virulence of urinary and bloodstream isolates of ExPEC *E. coli* in *G. mellonella,* it was found that even though ExPEC strains had a wide range of virulence factors, only a few of these factors were linked to a higher death rate and/or a strong immune response. Even though their genes

were only found in 10% of strains, *Afa/Dr* adhesins and *ompT* were linked to a high number of deaths in the *G. mellonella* model. Strains that had *ompT* and P-fimbriae made a lot more melanin than strains that did not have these surface structures (Ciesielczuk et al. 2015b).

This could explain the results of the little difference and refer us to a question of how stable CTX-M-15 is! However, the location of the chromosomal gene increases the stability of the strain/resistance combination, which will assure the vertical descent. For example, it will likely be much more stable for *ISEcP-1*-mediated insertions than the alternative, which involves having multiple copies of an identical sequence like *IS26*, which causes genetic instability and can promote homologous recombination events that result in the loss of all sequence (In this case, all resistance mechanisms) found between them, leaving a single copy of the element (Toleman and Walsh 2011).

In addition, despite our attempts to select strains that have near identical background genetics, there are other differences between these strains, such as the number of plasmids carried and whether any damage has occurred to the strain by the position of the insertion of the antibiotic gene in the bacterial chromosome. For example, we have demonstrated in Chapter three that insertion into the chromosome often destroys gene function by truncating a chromosomal gene. Such events may be more counterproductive in terms of fitness to cause an infection than the simultaneous loss of the plasmid causing an increased growth rate. In addition, there are also likely confounding limitations to the actual *G. mellonella* model at the present time. It is possible for there to be quantitative variances in the results due to the absence of universal genotypes of the larvae as well as the circumstances that have been implemented by each study group for the feeding, reproduction, and

preservation of the animals (Mowlds and Kavanagh 2008; Olsen et al. 2011; Loh et al. 2013). To substantiate these findings, a more extensive investigation of these contributing components is needed.

Chapter 7 General Discussion

Globally, antibiotic resistance is increasing to alarmingly high levels. Emerging and expanding globally, new resistance mechanisms impair our ability to treat common infectious diseases. As antibiotics become less effective, a growing number of infections, including bacteraemia and other infectious diseases, are becoming difficult and sometimes impossible to treat. The ESBL CTX-M-15 is arguably the most successful resistance mechanism that has emerged in *E. coli* since the start of the antibiotic era. The $bla_{\text{CTX-M-15}}$ gene was often found in isolates with chromosomal mutations giving rise to quinolone resistance and with co-carriage of the $bla_{\text{OXA-1}}$ β-lactamase and aminoglycoside resistance mechanisms in these initial *E. coli* isolates (Walsh et al. 2007; Toleman et al. 2012). Previous studies, as described in this study, have shown that human sewage can be used as an accurate measure of community carriage of *E. coli*. The acquisition of a gene from an external source that imparts resistance by any means is a key tactic that bacteria use to overcome the effects of antimicrobial activity. The study of ecological and evolutionary genetic structure and resistance mechanisms in *E. coli* is mutually beneficial and could play a role in improving the management of UTIs in the community by assessing commonly carried resistance mechanisms in sewage and then guiding appropriate therapy.

This project studied a large collection of *E. coli* isolates from numerous relevant sources and diverse geographic locations within the UK and investigated the chromosomal insertion of ESBL genes from different sources to investigate the importance of chromosomal carriage of antibiotic resistance genes. In addition, to get a true picture of the *E. coli* associated with human carriage in 2019, we characterised Saudi Arabian and Kazakh isolates using a wide range of molecular

bacteriology and functional genomics techniques, and we investigated the phylogenetic relationships between isolates from these different nations.

The results reported in **Chapter 3** showed that chromosomal carriage was slightly more important in bacteraemia isolates than in faecal ones, which may be due to the selection imposed in the clinical environment or the effect of losing a resistance plasmid, which increases the growth rates and thus makes it easier for the organism to establish infection. This particular area of research has been thoroughly investigated in the laboratory with different strains of bacteria; it has received little attention for contemporary wild-type isolates that are routinely carried in the human gut and cause serious infections in the community and hospital environments. Our study of more than 100 isolates that successfully located insertion sites and co-transposed genes is the first study that we know of that has used a large contemporary collection of *E. coli* isolates from both community and hospital sources. We managed to successfully close several genomes and locate the exact insertion sites as well as the size of transposed sections of DNA in the vast majority of the isolates studied. This demonstrated for the UK that chromosomal carriage of resistance genes is hugely important and that it is also a reasonably common event in other nations as well (Kazakhstan and Saudi Arabia).

The chromosomal insertions were mostly mediated by the genetic vehicles IS*26* and IS*EcP-1* insertions. In the majority of cases, these insertions were found inside structural genes rather than in intergenic regions. The *ISEc*P-1 element is more commonly responsible for insertion events and is hugely successful because it needs only one copy of the element for transposition of any gene that it manages to move beside. The fact that largely duplicated sections are not found with this type of insertion means that this will leave a more stable structure after transposition and

make homologous recombination events removing the inserted DNA less likely. This is in stark contrast to IS*26,* which needs two copies of the element to bracket a useful gene before it can be mobilised and then leaves repeated identical sections of DNA, which can subsequently be removed by homologous recombination events. Our study also demonstrates that *ISEc*P-1 can move large sections of DNA, even if it appears to have a bias towards moving smaller sections.

Although we explained several hotspots for chromosomal insertions, some of which were found not only in the UK collection but also in the Kazakh collection, as in *htrE* and *motA*. There seems to also be a general random nature to the chromosomal insertions, with insertions randomly arranged around the entire chromosome as indicated in Chapter 3 (**Figure 3.10**). Some of the hotspots may be due to the amplification of individual strains, or perhaps to stable resistance carriage, but others may actually provide an advantage to the strain by having individual genes knocked out. The insertion in the pepT gene may be an example of this, and this could be an avenue for future research. Random mutagenesis is a common laboratory approach utilised to explore the function and importance of genes. This particular collection of isolates includes many different insertions and gene knockouts that have been selected in the wild-type environment, and further research of any advantages or disadvantages may be an interesting approach to study the effect of gene knockouts from a natural collection.

The striking results of many resistance genes being co-transferred with the *bla*$_{\text{CTX-M-15}}$ gene, forming extensive resistance gene islands on the chromosome in some of the isolates, were not anticipated. This is particularly concerning as many of these strains leave only the carbapenems as good treatment options. A significant observation was that the movement of *bla*$_{\text{CTX-M-15}}$ to the *E. coli* chromosome was

closely related to the loss of the donating plasmid, with nearly 90% of isolates that had CTX-M-15 on the chromosome then subsequently losing the donating plasmid. Only a small number of isolates possessed both chromosome and plasmid versions of *bla*$_{\text{CTX-M-15}}$. This is important evidence that the plasmid paradox is a very real phenomenon in natural populations of *E. coli* and that the combination of the *bla*$_{\text{CTX-M-15}}$ gene and the highly mobile *ISEcP1* element favours chromosomal movement with subsequent loss of donating plasmids, as has been demonstrated in several laboratory populations.

The scarcity of thorough studies of community carriage of *E. coli* and antibiotic resistant *E. coli* in Saudi Arabia and Kazakhstan gave this project, in **Chapter 4**, huge importance since it is the first study to examine *E. coli* carriage in Unaizah, Saudi Arabia, and Karaganda, Kazakhstan. The study of *E. coli* without any antibiotic selection, with selection with cefotaxime, and selection with ciprofloxacin gave us a unique perception of the phylogroup and MLST changes to *E. coli* that happen when antibiotics are used in the community. To our knowledge, this is the first study examining two different nations and *E. coli* carriage with and without antibiotic selection by selecting *E. coli* from the same source under different selection pressures. The low percentage of virulent *E. coli* observed in Saudi Arabia compared to Kazakhstan and the UK raises many questions about the relation between the high sepsis rates in the UK compared to Saudi Arabia. However, these results differ from those in a study by (Al-Agamy et al. 2014), which investigated the molecular characteristics of ESBL-EC in Riyadh. It was found that twenty of 31 ESBL-EC clinical isolates (64%) were identified as the ST131 clone. Also, the molecular analysis of ESBL genotypes among ExPEC isolates revealed that CTX-M genes were highly prevalent when compared to the other β-lactamases encoding genes

such as *bla*$_{TEM}$, *bla*$_{SHV}$, and *bla*$_{OXA}$, and that *bla*$_{CTX\text{-}M\text{-}15}$ was the most common CTX-M variation (Alqasim et al. 2020).

Unfortunately, there are few sepsis studies in Saudi Arabia to confirm our observation, but we take the view that the general *E. coli* sepsis rates are likely much less than those found in the UK. We have shown that antibiotic selection can dramatically alter the structure of *E. coli* phylotypes and MLST and quit dramatically change a population of *E. coli* towards pathogenic phylotypes. This replacement of non-virulent *E. coli* types with virulent *E. coli* deserves to be further investigated as it demonstrates that the usage of antibiotics would not only increase antibiotic resistance rates but also select more virulent *E. coli* phylotypes and thus likely drive-up sepsis rates. This is an interesting observation that should be investigated in other countries. Notably, there is a big difference between the effect that selection has in Kazakhstan to that observed in Saudi Arabia, and it would be good to repeat these experiments in the UK as the effects may be different in different regions and perhaps help us understand why sepsis rates are different in different parts of the UK.

In the UK, *E. coli* is responsible for more cases of sepsis than any other organism, as well as more urinary tract infections than any other organism. Data show that sepsis rates have doubled since the UK implemented mandatory surveillance of *E. coli* bacteraemia in 2011. This project found variable levels of *E. coli* carriage between different nations and in different regions of the UK. **In Chapter 5**, we employed different typing methodologies and examined *E. coli* communities in six different locations alongside the M4 corridor and using Bangladesh as a comparator since we have previously shown that Bangladesh as well as several South Asian nations have very low carriage of pathogenic phylotypes of *E. coli*. There was a stark

difference in the community carriage of pathogenic *E. coli* between the UK and Bangladesh. The prevalence of pathogenic phylotypes was much higher in the UK compared to Bangladesh. The rising rates of sepsis within the UK are likely a direct result of the increasingly high carriage of pathogenic sequence types of *E. coli* in the community, which adds a novel understanding of the current UK sepsis crisis. We take the view that increasing gut carriage of pathogenic *E. coli* types is the cause of the rising sepsis rates in the UK, and we have shown that areas in the UK are associated with higher levels of sepsis. For example, South Wales has higher levels of pathogenic *E. coli* as compared to England. Between the sites selected at the East and West ends of the M4, we discovered a statistical difference in the prevalence of the main pathogenic B2 phylotype, but not for phylotype D. These findings were also statistically relevant for comparison between sites in Wales and England. Since South Asia consistently has low rates of pathogenic ExPEC *E. coli* types, we also collected sewage from Bangladesh during the same week as a comparison. We noted that the absence of pathogenic B2 types, which is consistent with previous findings for a number of S. Asian countries, including Pakistan, Bangladesh, and India. However, these results contradict with the (Stoppe et al. 2017) study, which indicates that there were no phylogroup sub structuring patterns and no association between the distribution of phylogroups and geographical location, climate, living area, eating habits, or date of collection. The statistical analysis that was done on our research made it very apparent that the phylotype composition of a population differs not only between countries but also within countries.

Also, our results indicate that the presence of just a few *E. coli* ST has a significant impact on sepsis rates. Our analysis demonstrates that several varieties of ST131

are carried in the UK, including clades that have been differentiated by fimH alleles. We identified all these clades and reaffirmed that the *bla*$_{\text{CTX-M-15}}$ gene is most frequently linked to the *fimH30* allele. It is interesting to note that *bla*$_{\text{CTX-M-27}}$ was just as prevalent in our collection as *bla*$_{\text{CTX-M-15}}$, which may indicate a change in the dominant ESBL since (Day et al. 2019) previous study. Importantly, we discovered in this study that other pathogenic types are not strongly correlated with antibiotic resistance; only ST131 and ST69 are. Also, the finding that ST127 has a very high prevalence in our collection, particularly at the westerly locations, suggests that this MLST is crucial to the UK sepsis and that its prevalence has been rising recently. The importance of this MLST, which has only recently been described in the UK, is dramatically highlighted by this study.

The use of the *Galleria mellonella* infection model, described in **Chapter 6,** to investigate pathogenic *E. coli* has significantly increased in the last decade. We sought to investigate the connection between the genetic location of *bla*$_{\text{CTX-M-15}}$ in *E. coli* ST131 and virulence using this model. There were slight differences in mortality rates between the *E. coli* isolates carrying *bla*$_{\text{CTX-M-15}}$ in the chromosome only, the plasmid only, and the ones carrying this resistance gene in both the chromosome and the plasmid. We selected strains that have near identical background genetics (all ST131 with the *fimH30* allele, i.e., essential clonal). However, there are other differences between these strains, such as the number of plasmids carried and whether any damage has occurred to the strain and the position of the insertion of the antibiotic gene in the bacterial chromosome. However, the comparison between growth rates of strains having a chromosomal only position as compared to the two groups containing plasmid locations indicated that strains with a sole chromosomal location grew slightly faster. The strains with their resistance genes found on the

chromosome grew faster in the growth experiment than those belonging to the other two groups. However, this observation was not replicated in the virulence assays, as there was no statistical difference between the groups. We think that this result is somewhat hidden by the differences in the number of plasmids and insertion sites, while these experiments could be improved in the future by ensuring that comparator strains have identical background genotypes with regard to plasmid numbers and genes knocked out.

7.1 Key strengths and limitations of the thesis

The study of more than 100 isolates and successfully locating the chromosomal insertions and managing to close several genomes allowed us to observe that movement to the chromosome is a common occurrence in *E. coli*. This has not been done before with such a comprehensive *E. coli* collection from several sources and with such a depth of analysis. Another significant observation was that the movement of $bla_{\text{CTX-M-15}}$ to the *E. coli* chromosome was closely related to the loss of the donating plasmid, as suggested by the plasmid paradox and several laboratory experiments, adding validation to prior laboratory work. Surprisingly, we noted that several cell lines had lost the $bla_{\text{CTX-M-15}}$ gene entirely, suggesting perhaps that IS*Ecp1* is dangerous to harbour and that movement events that are common with this element can lead to catastrophic events for the *E. coli* strain.

The scarcity of thorough studies in Saudi Arabia and Kazakhstan gave this project huge importance since it is the first study to examine *E. coli* carriage in Unaizah, Saudi Arabia, and Karaganda, Kazakhstan. The study of *E. coli* without any antibiotic, selective with cefotaxime, and selective with ciprofloxacin gave us a unique perception of phylogroup, MLST, and antibiotic resistance. To the best of our knowledge, this is the first study to look at *E. coli* carriage in three different countries. However, only a few isolates that were selected with ciprofloxacin were whole genome sequenced due to COVID circumstances.

The use of different typing methodologies and the examination of *E. coli* communities in six different sites in the UK. The carriage of pathogenic *E. coli* correlating with high sepsis rates adds a novel understanding of the current UK sepsis crisis by showing that areas in the UK are associated with higher levels of sepsis, which directly contradicts the assumptions of prior Brazilian publications.

Also, our results indicate that a few *E. coli* STs have a significant impact on sepsis rates, specifically ST131. Moreover, we observed that ST131 and ST69 are strongly correlated with antibiotic resistance, more than other pathogenic types. Lastly, the finding that ST127 has a very high prevalence in our collection, which correlates with other studies, shows the importance of ST127.

In retrospect, I should have raised the number of isolates in Chapter 4 and added one site from the UK to observe the differences in the human carriage of *E. coli*. Due to the COVID-19 pandemic, I lacked access to the labs, which reduced my ability to identify new isolates and sequence them, though we had done several collections within Saudi Arabia. I should have collected sepsis isolates from Saudi Arabia to compare with the UK collection, and that would have given us a great opportunity to observe the differences between these countries. Unfortunately, this was not an easy thing to accomplish with various permissions and with the COVID world shutdown.

7.2 Future work recommendations from the thesis

This thesis presents a guide and application of methods that may be used to enhance the evidence base for the AMR-associated burden in the antibiotic resistance of *E. coli* and other microbes.

As a Saudi researcher, I can use this thesis for further research in applying the method to study sepsis in the Arabian Peninsula. Saudi Arabia has swiftly risen to become one of the most prominent participants in Middle Eastern scientific research, which will give me a huge opportunity to study the burden of antibiotic resistance.

7.3 Final conclusion

The results of this study show that antibiotic resistance is a complicated and diverse problem that poses a serious threat to public health. This highlighted the importance of implementing a surveillance scheme for *E. coli* carriage that aims to reduce *E. coli* infections. Therefore, it is recommended to use an untreated sewage sample on entry as a surveillance tool to assay antibiotic resistance in *E. coli*. The chromosomal carriage of antibiotic resistance is hugely important and could be used to look at the importance of hot spots of insertion and the stability of chromosomal carriage worldwide. The project is likely to help us understand differences in global resistance rates and perhaps identify intervention routes. Moreover, the project contains low risk approaches that will generate a lot of data together with research angles and questions that will present opportunities to conceive design and develop research ability and integrity. The methodologies described throughout this thesis can be applied to different microorganisms and various environments to help narrow the present knowledge gap on the impact of AMR.

Chapter 8 References

Abd El Ghany, M., Sharaf, H., Al-Agamy, M.H., Shibl, A., Hill-Cawthorne, G.A. and Hong, P.-Y. 2018. Genomic characterization of NDM-1 and 5, and OXA-181 carbapenemases in uropathogenic Escherichia coli isolates from Riyadh, Saudi Arabia. *PloS one* 13(8), p. e0201613. doi: 10.1371/journal.pone.0201613.

Abernethy, J. et al. 2017. Epidemiology of Escherichia coli bacteraemia in England: results of an enhanced sentinel surveillance programme. *The Journal of hospital infection* 95(4), pp. 365–375. doi: 10.1016/j.jhin.2016.12.008.

Al-Agamy, M.H., Aljallal, A., Radwan, H.H. and Shibl, A.M. 2018. Characterization of carbapenemases, ESBLs, and plasmid-mediated quinolone determinants in carbapenem-insensitive Escherichia coli and Klebsiella pneumoniae in Riyadh hospitals. *Journal of Infection and Public Health* 11(1), pp. 64–68. doi: 10.1016/j.jiph.2017.03.010.

Al-Agamy, M.H., Shibl, A.M., Hafez, M.M., Al-Ahdal, M.N., Memish, Z.A. and Khubnani, H. 2014. Molecular characteristics of extended-spectrum β-lactamase-producing Escherichia coli in Riyadh: emergence of CTX-M-15-producing E. coli ST131. *Annals of clinical microbiology and antimicrobials* 13, p. 4. doi: 10.1186/1476-0711-13-4.

Alghamdi, S. et al. 2021. Antimicrobial Stewardship Programmes in Saudi Hospitals: Evidence from a National Survey. *Antibiotics (Basel, Switzerland)* 10(2). doi: 10.3390/antibiotics10020193.

Alghoribi, M.F., Gibreel, T.M., Dodgson, A.R., Beatson, S.A. and Upton, M. 2014. Galleria mellonella infection model demonstrates high lethality of ST69 and ST127 uropathogenic E. coli. *PLoS ONE* 9(7). doi: 10.1371/journal.pone.0101547.

Alghoribi, M.F., Gibreel, T.M., Farnham, G., Al Johani, S.M., Balkhy, H.H. and Upton, M. 2015. Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic Escherichia coli

clones in Riyadh, Saudi Arabia. *Journal of Antimicrobial Chemotherapy* 70(10), pp. 2757–2762. doi: 10.1093/jac/dkv188.

Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. 2016. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR Escherichia coli ST73: High levels of genomic and plasmid diversity in causative isolates. *Journal of Antimicrobial Chemotherapy* 71(2), pp. 339–343. doi: 10.1093/jac/dkv365.

Alqasim, A., Abu Jaffal, A. and Alyousef, A.A. 2018. Prevalence of Multidrug Resistance and Extended-Spectrum β-Lactamase Carriage of Clinical Uropathogenic Escherichia coli Isolates in Riyadh, Saudi Arabia. *International journal of microbiology* 2018, p. 3026851. doi: 10.1155/2018/3026851.

Alqasim, A., Abu Jaffal, A. and Alyousef, A.A. 2020. Prevalence and molecular characteristics of sequence type 131 clone among clinical uropathogenic Escherichia coli isolates in Riyadh, Saudi Arabia. *Saudi journal of biological sciences* 27(1), pp. 296–302. doi: 10.1016/j.sjbs.2019.09.020.

Alwash, M.S. and Al-Rafyai, H.M. 2019. Antibiotic resistance patterns of diverse Escherichia coli phylogenetic groups isolated from the Al-Hillah river in Babylon province, Iraq. *Scientific World Journal* 2019. doi: 10.1155/2019/5927059.

Alyamani, E.J., Khiyami, A.M., Booq, R.Y., Majrashi, M.A., Bahwerth, F.S. and Rechkina, E. 2017. The occurrence of ESBL-producing Escherichia coli carrying aminoglycoside resistance genes in urinary tract infections in Saudi Arabia. *Annals of Clinical Microbiology and Antimicrobials* 16(1), p. 1. doi: 10.1186/s12941-016-0177-6.

Andrews, R.H. and Chilton, N.B. 1999. Multilocus enzyme electrophoresis: a valuable technique for providing answers to problems in parasite systematics. *International journal for parasitology* 29(2), pp. 213–253. doi: 10.1016/s0020-7519(98)00168-4.

Antoine, C. et al. 2021. In Vitro Characterization and In Vivo Efficacy Assessment in Galleria mellonella Larvae of Newly Isolated Bacteriophages against Escherichia coli K1. *Viruses* 13(10). doi: 10.3390/v13102005.

Arthur, A., Nimmo, E., Hettle, S. and Sherratt, D. 1984. Transposition and transposition immunity of transposon Tn3 derivatives having different ends. *The EMBO journal* 3(8), pp. 1723–1729. doi: 10.1002/j.1460-2075.1984.tb02038.x.

Avila, P., de la Cruz, F., Ward, E. and Grinsted, J. 1984. Plasmids containing one inverted repeat of Tn21 can fuse with other plasmids in the presence of Tn21 transposase. *Molecular & general genetics : MGG* 195(1–2), pp. 288–293. doi: 10.1007/BF00332761.

Babenko, D., Sands, K., Portal, E., Walsh, T.R. and Toleman, M.A. 2019. Early emergence and dissemination of Escherichia coli ST131 H30 Rx from India. *European conference of clinical Microbiology and Infectious diseases*, p. P1453.

Bailey, J.K., Pinyon, J.L., Anantham, S. and Hall, R.M. 2010. Commensal Escherichia coli of healthy humans: a reservoir for antibiotic-resistance determinants. *Journal of medical microbiology* 59(Pt 11), pp. 1331–1339. doi: 10.1099/jmm.0.022475-0.

Banerjee, R. and Johnson, J.R. 2014. A new clone sweeps clean: the enigmatic emergence of Escherichia coli sequence type 131. *Antimicrobial agents and chemotherapy* 58(9), pp. 4997–5004. doi: 10.1128/AAC.02824-14.

Batchelor, M., Threlfall, E.J. and Liebana, E. 2005. Cephalosporin resistance among animal-associated Enterobacteria: a current perspective. *Expert Review of Anti-infective Therapy* 3(3), pp. 403–417. Available at: https://doi.org/10.1586/14787210.3.3.403.

Bidla, G., Hauling, T., Dushay, M.S. and Theopold, U. 2009. Activation of insect phenoloxidase after injury: endogenous versus foreign elicitors. *Journal of innate immunity* 1(4), pp. 301–308. doi: 10.1159/000168009.

Bonnet, R. et al. 2002. Chromosome-encoded class D beta-lactamase OXA-23 in Proteus mirabilis. *Antimicrobial agents and chemotherapy* 46(6), pp. 2004–2006. doi: 10.1128/AAC.46.6.2004-2006.2002.

Bourne, J.A., Chong, W.L. and Gordon, D.M. 2019. Genetic structure, antimicrobial resistance and frequency of human associated Escherichia coli sequence types among faecal isolates from healthy dogs and cats living in Canberra, Australia. *PloS one* 14(3), p. e0212867. doi: 10.1371/journal.pone.0212867.

Brenner, D.J., Fanning, G.R., Steigerwalt, A.G., Orskov, I. and Orskov, F. 1972. Polynucleotide sequence relatedness among three groups of pathogenic Escherichia coli strains. *Infection and immunity* 6(3), pp. 308–315. doi: 10.1128/iai.6.3.308-315.1972.

Brockhurst, M.A. and Harrison, E. 2022. Ecological and evolutionary solutions to the plasmid paradox. *Trends in microbiology* 30(6), pp. 534–543. doi: 10.1016/j.tim.2021.11.001.

Broda, P., E. 1979. Plasmids. *W.H. Freman & Co., Oxford.*

Burrus, V., Pavlovic, G., Decaris, B. and Guédon, G. 2002. Conjugative transposons: the tip of the iceberg. *Molecular Microbiology* 46(3), pp. 601–610. doi: https://doi.org/10.1046/j.1365-2958.2002.03191.x.

Bush, K. and Jacoby, G.A. 2010a. Updated functional classification of β-lactamases. *Antimicrobial Agents and Chemotherapy* 54(3), pp. 969–976. doi: 10.1128/AAC.01009-09.

Bush, K. and Jacoby, G.A. 2010b. Updated functional classification of β-lactamases. *Antimicrobial Agents and Chemotherapy* 54(3), pp. 969–976. doi: 10.1128/AAC.01009-09.

Bush, K., Jacoby, G.A. and Medeiros, A.A. 1995. A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrobial agents and chemotherapy* 39(6), pp. 1211–1233. doi: 10.1128/AAC.39.6.1211.

Cantón, R. and Coque, T.M. 2006. The CTX-M β-lactamase pandemic. *Current Opinion in Microbiology* 9(5), pp. 466–475. doi: https://doi.org/10.1016/j.mib.2006.08.011.

Cantón, R., González-Alba, J.M. and Galán, J.C. 2012. CTX-M Enzymes: Origin and Diffusion. *Frontiers in microbiology* 3, p. 110. doi: 10.3389/fmicb.2012.00110.

Carlos, C. et al. 2010. Escherichia coli phylogenetic group determination and its application in the identification of the major animal source of fecal contamination. *BMC microbiology* 10, p. 161. doi: 10.1186/1471-2180-10-161.

CDC. 2024. *STANDARD OPERATING PROCEDURE FOR PULSENET PFGE OF ESCHERICHIA COLI O157:H7, ESCHERICHIA COLI NON-O157 (STEC), SALMONELLA SEROTYPES, SHIGELLA SONNEI AND SHIGELLA FLEXNERI*.

Chattaway, M.A., Schaefer, U., Tewolde, R., Dallman, T.J. and Jenkins, C. 2017. Identification of Escherichia coli and Shigella Species from Whole-Genome Sequences. *Journal of clinical microbiology* 55(2), pp. 616–623. doi: 10.1128/JCM.01790-16.

Chopra, I. and Greenwood, D. 2001. Antibacterial Agents: Basis of Action. In: *eLS*. doi: https://doi.org/10.1038/npg.els.0001992.

Ciesielczuk, H., Betts, J., Phee, L., Doumith, M., Hope, R., Woodford, N. and Wareham, D.W. 2015a. Comparative virulence of urinary and bloodstream isolates of extra-intestinal pathogenic Escherichia coli in a Galleria mellonella model. *Virulence* 6(2), pp. 145–151. doi: 10.4161/21505594.2014.988095.

Ciesielczuk, H., Betts, J., Phee, L., Doumith, M., Hope, R., Woodford, N. and Wareham, D.W. 2015b. Comparative virulence of urinary and bloodstream isolates of extra-intestinal pathogenic Escherichia coli in a Galleria mellonella model. *Virulence* 6(2), pp. 145–151. doi: 10.4161/21505594.2014.988095.

Clermont, O. et al. 2008. The CTX-M-15-producing Escherichia coli diffusing clone belongs to a highly virulent B2 phylogenetic subgroup. *The Journal of antimicrobial chemotherapy* 61(5), pp. 1024–1028. doi: 10.1093/jac/dkn084.

Clermont, O. et al. 2009. Rapid detection of the O25b-ST131 clone of Escherichia coli encompassing the CTX-M-15-producing strains. *The Journal of antimicrobial chemotherapy* 64(2), pp. 274–277. doi: 10.1093/jac/dkp194.

Clermont, O., Bonacorsi, S. and Bingen, E. 2000. Rapid and simple determination of the Escherichia coli phylogenetic group. *Applied and Environmental Microbiology* 66(10), pp. 4555–4558. doi: 10.1128/AEM.66.10.4555-4558.2000.

Clermont, O., Christenson, J.K., Denamur, E. and Gordon, D.M. 2013. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports* 5(1), pp. 58–65. Available at: https://doi.org/10.1111/1758-2229.12019.

Codjoe, F.S. and Donkor, E.S. 2017. Carbapenem Resistance: A Review. *Medical sciences (Basel, Switzerland)* 6(1). doi: 10.3390/medsci6010001.

Common, I.F.B. 1975. Evolution and Classification of the Lepidoptera. *Annual Review of Entomology* 20(1), pp. 183–203. doi: 10.1146/annurev.en.20.010175.001151.

Conway, T. and Cohen, P.S. 2015. Commensal and Pathogenic Escherichia coli Metabolism in the Gut. *Microbiology spectrum* 3(3). doi: 10.1128/microbiolspec.MBP-0006-2014.

Coque, T.M. et al. 2008. Dissemination of clonally related Escherichia coli strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerging infectious diseases* 14(2), pp. 195–200. doi: 10.3201/eid1402.070350.

Coura, F.M., Diniz, S. de A., Silva, M.X., Mussi, J.M.S., Barbosa, S.M., Lage, A.P. and Heinemann, M.B. 2015. Phylogenetic Group Determination of Escherichia coli Isolated from Animals Samples. *TheScientificWorldJournal* 2015, p. 258424. doi: 10.1155/2015/258424.

Da Cunha, V. et al. 2014. Streptococcus agalactiae clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nature Communications* 5, pp. 1–12. doi: 10.1038/ncomms5544.

Damborg, P., Nielsen, S.S. and Guardabassi, L. 2009. Escherichia coli shedding patterns in humans and dogs: insights into within-household transmission of phylotypes associated with urinary tract infections. *Epidemiology and infection* 137(10), pp. 1457–1464. doi: 10.1017/S095026880900226X.

D'Andrea, M.M. et al. 2011. Evolution and spread of a multidrug-resistant Proteus mirabilis clone with chromosomal AmpC-type cephalosporinases in Europe. *Antimicrobial agents and chemotherapy* 55(6), pp. 2735–2742. doi: 10.1128/AAC.01736-10.

Daoud, Z. and Afif, C. 2011. Escherichia coli Isolated from Urinary Tract Infections of Lebanese Patients between 2000 and 2009: Epidemiology and Profiles of Resistance. *Chemotherapy research and practice* 2011, p. 218431. doi: 10.1155/2011/218431.

Davies, J. and Davies, D. 2010. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR* 74(3), pp. 417–433. doi: 10.1128/MMBR.00016-10.

Day, M.J. et al. 2019. Extended-spectrum β-lactamase-producing Escherichia coli in human-derived and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. *The Lancet. Infectious diseases* 19(12), pp. 1325–1335. doi: 10.1016/S1473-3099(19)30273-7.

Decano, A.G., Ludden, C., Feltwell, T., Judge, K., Parkhill, J. and Downing, T. 2019. Complete Assembly of Escherichia coli Sequence Type 131 Genomes Using Long Reads Demonstrates

Antibiotic Resistance Gene Variation within Diverse Plasmid and Chromosomal Contexts. *mSphere* 4(3). doi: 10.1128/mSphere.00130-19.

Demir, D., Gençer, N. and Er, A. 2012. Purification and characterization of prophenoloxidase from Galleria mellonella L. *Artificial cells, blood substitutes, and immobilization biotechnology* 40(6), pp. 391–395. doi: 10.3109/10731199.2012.696060.

Denamur, E., Clermont, O., Bonacorsi, S. and Gordon, D. 2021. The population genetics of pathogenic Escherichia coli. *Nature Reviews Microbiology* 19(1), pp. 37–54. Available at: https://doi.org/10.1038/s41579-020-0416-x.

Desalermos, A., Fuchs, B.B. and Mylonakis, E. 2012. Selecting an invertebrate model host for the study of fungal pathogenesis. *PLoS pathogens* 8(2), p. e1002451. doi: 10.1371/journal.ppat.1002451.

Doumith, M. et al. 2015. Rapid identification of major Escherichia coli sequence types causing urinary tract and bloodstream infections. *Journal of Clinical Microbiology* 53(1), pp. 160–166. doi: 10.1128/JCM.02562-14.

Doumith, M., Day, M.J., Hope, R., Wain, J. and Woodford, N. 2012. Improved multiplex PCR strategy for rapid assignment of the four major Escherichia coli phylogenetic groups. *Journal of Clinical Microbiology* 50(9), pp. 3108–3110. doi: 10.1128/JCM.01468-12.

Dubnau, D. 1999. DNA uptake in bacteria. *Annual review of microbiology* 53, pp. 217–244. doi: 10.1146/annurev.micro.53.1.217.

Duriez, P. et al. 2001a. Commensal Escherichia coli isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology (Reading, England)* 147(Pt 6), pp. 1671–1676. doi: 10.1099/00221287-147-6-1671.

Duriez, P. et al. 2001b. Commensal Escherichia coli isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology* 147(6), pp. 1671–1676. doi: 10.1099/00221287-147-6-1671.

Dykhuizen, D.E. and Dean, A.M. 1990. Enzyme activity and fitness: Evolution in solution. *Trends in ecology & evolution* 5(8), pp. 257–262. doi: 10.1016/0169-5347(90)90067-N.

Erol, H.B., Kaskatepe, B., Ozturk, S. and Safi Oz, Z. 2022. The comparison of lytic activity of isolated phage and commercial Intesti bacteriophage on ESBL producer E. coli and determination of Ec_P6 phage efficacy with in vivo Galleria mellonella larvae model. *Microbial pathogenesis* 167, p. 105563. doi: 10.1016/j.micpath.2022.105563.

Escobar-Páramo, P. et al. 2004. Large-scale population structure of human commensal Escherichia coli isolates. *Applied and Environmental Microbiology* 70(9), pp. 5698–5700. doi: 10.1128/AEM.70.9.5698-5700.2004.

Etebu, E. and Arikekpar, I. 2016. Antibiotics: Classification and mechanisms of action with emphasis on molecular perspectives. *Int. J. Appl. Microbiol. Biotechnol. Res* 4(2016), pp. 90–101.

Excellence., N.I. for H. and C. 2018. *Urinary tract infection (lower) - men, Prevalence.* p. Clinical knowledge summary.

Excellence., N.I. for H. and C. 2021. *Urinary tract infection (lower) - women, Prevalence.* p. Clinical knowledge summary.

Fedhila, S. et al. 2010. Comparative analysis of the virulence of invertebrate and mammalian pathogenic bacteria in the oral insect infection model Galleria mellonella. *Journal of invertebrate pathology* 103(1), pp. 24–29. doi: 10.1016/j.jip.2009.09.005.

Fehér, T. et al. 2012. Competition between transposable elements and mutator genes in bacteria. *Molecular biology and evolution* 29(10), pp. 3153–3159. doi: 10.1093/molbev/mss122.

Flores-Mireles, A.L., Walker, J.N., Caparon, M. and Hultgren, S.J. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nature reviews. Microbiology* 13(5), pp. 269–284. doi: 10.1038/nrmicro3432.

Foxman, B. 2002. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *The American journal of medicine* 113 Suppl, pp. 5S-13S. doi: 10.1016/s0002-9343(02)01054-9.

Fuchs, B.B., O'Brien, E., Khoury, J.B.E. and Mylonakis, E. 2010. Methods for using Galleria mellonella as a model host to study fungal pathogenesis. *Virulence* 1(6), pp. 475–482. doi: 10.4161/viru.1.6.12985.

Gerhard Domagk – Facts - NobelPrize.org. [no date]. Available at: https://www.nobelprize.org/prizes/medicine/1939/domagk/facts/ [Accessed: 24 February 2024].

Giannouli, M. et al. 2014. Use of larvae of the wax moth Galleria mellonella as an in vivo model to study the virulence of Helicobacter pylori. *BMC microbiology* 14, p. 228. doi: 10.1186/s12866-014-0228-0.

Gibreel, T.M., Dodgson, A.R., Cheesbrough, J., Fox, A.J., Bolton, F.J. and Upton, M. 2012. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. *Journal of Antimicrobial Chemotherapy* 67(2), pp. 346–356. Available at: https://doi.org/10.1093/jac/dkr451.

Gordon, D.M. and Cowling, A. 2003. The distribution and genetic structure of Escherichia coli in Australian vertebrates: Host and geographic effects. *Microbiology* 149(12), pp. 3575–3586. doi: 10.1099/mic.0.26486-0.

Goswami, C., Fox, S., Holden, M.T.G., Connor, M., Leanord, A. and Evans, T.J. 2020. Origin, maintenance and spread of antibiotic resistance genes within plasmids and chromosomes of bloodstream isolates of Escherichia coli. *Microbial genomics* 6(4). doi: 10.1099/mgen.0.000353.

Guzmán, M.C., Bistoni, M. de los A., Tamagnini, L.M. and González, R.D. 2004. Recovery of Escherichia coli in fresh water fish, Jenynsia multidentata and Bryconamericus iheringi. *Water research* 38(9), pp. 2367–2373. doi: 10.1016/j.watres.2004.02.016.

Hall, B.G. 1978. Experimental evolution of a new enzymatic function. II. Evolution of multiple functions for ebg enzyme in E. coli. *Genetics* 89(3), pp. 453–465. doi: 10.1093/genetics/89.3.453.

Harada, S., Ishii, Y., Saga, T., Kouyama, Y., Tateda, K. and Yamaguchi, K. 2012. Chromosomal integration and location on IncT plasmids of the blaCTX-M-2 gene in Proteus mirabilis clinical isolates. *Antimicrobial agents and chemotherapy* 56(2), pp. 1093–1096. doi: 10.1128/AAC.00258-11.

Harding, C.R., Schroeder, G.N., Collins, J.W. and Frankel, G. 2013. Use of Galleria mellonella as a model organism to study Legionella pneumophila infection. *Journal of visualized experiments : JoVE* (81), p. e50964. doi: 10.3791/50964.

Harrison, E. and Brockhurst, M.A. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in microbiology* 20(6), pp. 262–267. doi: 10.1016/j.tim.2012.04.003.

Heitmueller, M., Billion, A., Dobrindt, U., Vilcinskas, A. and Mukherjee, K. 2017. Epigenetic Mechanisms Regulate Innate Immunity against Uropathogenic and Commensal-Like Escherichia coli in the Surrogate Insect Model Galleria mellonella. *Infection and immunity* 85(10). doi: 10.1128/IAI.00336-17.

Hendrickson, H. 2009. Order and disorder during Escherichia coli divergence. *PLoS Genetics* 5(1), pp. 1–2. doi: 10.1371/journal.pgen.1000335.

Hirai, I., Fukui, N., Taguchi, M., Yamauchi, K., Nakamura, T., Okano, S. and Yamamoto, Y. 2013. Detection of chromosomal blaCTX-M-15 in Escherichia coli O25b-B2-ST131 isolates from the Kinki region of Japan. *International journal of antimicrobial agents* 42(6), pp. 500–506. doi: 10.1016/j.ijantimicag.2013.08.005.

Hobman, J.L., Penn, C.W. and Pallen, M.J. 2007. Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully? *Molecular microbiology* 64(4), pp. 881–885. doi: 10.1111/j.1365-2958.2007.05710.x.

Imai, K. et al. 2019. Clinical characteristics in blood stream infections caused by Klebsiella pneumoniae, Klebsiella variicola, and Klebsiella quasipneumoniae: a comparative study, Japan, 2014-2017. *BMC infectious diseases* 19(1), p. 946. doi: 10.1186/s12879-019-4498-x.

Ishii, S., Ksoll, W.B., Hicks, R.E. and Sadowsky, M.J. 2006. Presence and growth of naturalized Escherichia coli in temperate soils from Lake Superior watersheds. *Applied and environmental microbiology* 72(1), pp. 612–621. doi: 10.1128/AEM.72.1.612-621.2006.

Jakobsen, L. et al. 2010. Escherichia coli isolates from broiler chicken meat, broiler chickens, pork, and pigs share phylogroups and antimicrobial resistance with community-dwelling humans and patients with urinary tract infection. *Foodborne pathogens and disease* 7(5), pp. 537–547. doi: 10.1089/fpd.2009.0409.

Jaureguy, F. et al. 2008. Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. *BMC genomics* 9, p. 560. doi: 10.1186/1471-2164-9-560.

Jemel, S., Guillot, J., Kallel, K., Botterel, F. and Dannaoui, E. 2020. Galleria mellonella for the evaluation of antifungal efficacy against medically important fungi, a narrative review. *Microorganisms* 8(3). doi: 10.3390/microorganisms8030390.

Joensen, K.G., Tetzschner, A.M.M., Iguchi, A., Aarestrup, F.M. and Scheutz, F. 2015. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *Journal of clinical microbiology* 53(8), pp. 2410–2426. doi: 10.1128/JCM.00008-15.

Johnson, J.R. et al. 2013. Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. *The Journal of infectious diseases* 207(6), pp. 919–928. doi: 10.1093/infdis/jis933.

Johnson, J.R., Johnston, B.D., Porter, S., Thuras, P., Aziz, M. and Price, L.B. 2019. Accessory Traits and Phylogenetic Background Predict Escherichia coli Extraintestinal Virulence Better Than Does Ecological Source. *The Journal of infectious diseases* 219(1), pp. 121–132. doi: 10.1093/infdis/jiy459.

Kader, A.A., Kumar, A. and Dass, S.M. 2004. Antimicrobial resistance patterns of gram-negative bacteria isolated from urine cultures at a general hospital. *Saudi journal of kidney diseases and transplantation : an official publication of the Saudi Center for Organ Transplantation, Saudi Arabia* 15(2), pp. 135–139.

Kapoor, G., Saigal, S. and Elongavan, A. 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. Available at: www.joacp.org [Accessed: 22 February 2024].

Karim, A., Poirel, L., Nagarajan, S. and Nordmann, P. 2001. Plasmid-mediated extended-spectrum beta-lactamase (CTX-M-3 like) from India and gene association with insertion sequence ISEcp1. *FEMS microbiology letters* 201(2), pp. 237–241. doi: 10.1111/j.1574-6968.2001.tb10762.x.

Karisik, E., Ellington, M.J., Pike, R., Warren, R.E., Livermore, D.M. and Woodford, N. 2006. Molecular characterization of plasmids encoding CTX-M-15 beta-lactamases from Escherichia coli strains in the United Kingdom. *The Journal of antimicrobial chemotherapy* 58(3), pp. 665–668. doi: 10.1093/jac/dkl309.

Karl, D., Muhammad, M., J., K.R. and Xilin, Z. 2008. Quinolone-Mediated Bacterial Death. *Antimicrobial Agents and Chemotherapy* 52(2), pp. 385–392. Available at: https://doi.org/10.1128/aac.01617-06.

Kauffmann, F. 1947. The serology of the coli group. *Journal of immunology (Baltimore, Md. : 1950)*.

Kohanski, M.A., Dwyer, D.J. and Collins, J.J. 2010. How antibiotics kill bacteria: from targets to networks. *Nature reviews. Microbiology* 8(6), p. 423. Available at: /pmc/articles/PMC2896384/ [Accessed: 24 February 2024].

Land, M. et al. 2015. Insights from 20 years of bacterial genome sequencing. *Functional & integrative genomics* 15(2), pp. 141–161. doi: 10.1007/s10142-015-0433-4.

Lapage, S.P. 1979. Proposal of Enterobacteraceae nom . nov . as a Substitute for the Illegitimate but Conserved Name Enterobacteriaceae Rahn 1937 Request for an Opinion. 1960(July), pp. 265–266.

Lee, Y.S. et al. 2004. Purification, cDNA cloning and expression of an insect defensin from the great wax moth, Galleria mellonella. *Insect molecular biology* 13(1), pp. 65–72. doi: 10.1111/j.1365-2583.2004.00462.x.

Leuko, S. and Raivio, T.L. 2012. Mutations that impact the enteropathogenic escherichia coli Cpx envelope stress response attenuate virulence in galleria mellonella. *Infection and Immunity* 80(9), pp. 3077–3085. doi: 10.1128/IAI.00081-12.

Li, B., Sun, J.Y., Han, L.Z., Huang, X.H., Fu, Q. and Ni, Y.X. 2010. Phylogenetic groups and pathogenicity island markers in fecal escherichia coli isolates from asymptomatic humans in china. *Applied and Environmental Microbiology* 76(19), pp. 6698–6700. doi: 10.1128/AEM.00707-10.

Li, D., Scherfer, C., Korayem, A.M., Zhao, Z., Schmidt, O. and Theopold, U. 2002. Insect hemolymph clotting: evidence for interaction between the coagulation system and the prophenoloxidase activating cascade. *Insect biochemistry and molecular biology* 32(8), pp. 919–928. doi: 10.1016/s0965-1748(02)00030-9.

Livermore, D.M. et al. 2007. CTX-M: changing the face of ESBLs in Europe. *The Journal of antimicrobial chemotherapy* 59(2), pp. 165–174. doi: 10.1093/jac/dkl483.

Loh, J.M.S., Adenwalla, N., Wiles, S. and Proft, T. 2013. Galleria mellonella larvae as an infection model for group A streptococcus. *Virulence* 4(5), pp. 419–428. doi: 10.4161/viru.24930.

Lukjancenko, O., Wassenaar, T.M. and Ussery, D.W. 2010. Comparison of 61 Sequenced Escherichia coli Genomes. *Microbial Ecology* 60(4), pp. 708–720. Available at: https://doi.org/10.1007/s00248-010-9717-3.

Luo, Y., Mao, D., Rysz, M., Zhou, Q., Zhang, H., Xu, L. and Alvarez, P.J.J. 2010. Trends in antibiotic resistance genes occurrence in the Haihe River, China. *Environmental Science and Technology* 44(19), pp. 7220–7225. doi: 10.1021/es100233w.

Mahillon, J. and Chandler, M. 1998. Insertion sequences. *Microbiology and molecular biology reviews : MMBR* 62(3), pp. 725–774. doi: 10.1128/MMBR.62.3.725-774.1998.

Maiden, M.C. et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* 95(6), pp. 3140–3145. doi: 10.1073/pnas.95.6.3140.

Maiden, M.C.J. 2006. Multilocus sequence typing of bacteria. *Annual review of microbiology* 60, pp. 561–588. doi: 10.1146/annurev.micro.59.030804.121325.

Martens, E. and Demain, A.L. 2017. The antibiotic resistance crisis, with a focus on the United States. *The Journal of antibiotics* 70(5), pp. 520–526. doi: 10.1038/ja.2017.30.

Mathai, D., Rhomberg, P.R., Biedenbach, D.J. and Jones, R.N. 2002. Evaluation of the in vitro activity of six broad-spectrum beta-lactam antimicrobial agents tested against recent clinical isolates from India: a survey of ten medical center laboratories. *Diagnostic microbiology and infectious disease* 44(4), pp. 367–377. doi: 10.1016/s0732-8893(02)00466-2.

Mazel, D., Dychinco, B., Webb, V.A. and Davies, J. 1998. A distinctive class of integron in the Vibrio cholerae genome. *Science (New York, N.Y.)* 280(5363), pp. 605–608. doi: 10.1126/science.280.5363.605.

Mbler, R.P.A. 1980. *The structure of (5-lactamases*.

Melzer, M. and Petersen, I. 2007. Mortality following bacteraemic infection caused by extended spectrum beta-lactamase (ESBL) producing E. coli compared to non-ESBL producing E. coli. *The Journal of infection* 55(3), pp. 254–259. doi: 10.1016/j.jinf.2007.04.007.

Ménard, G., Rouillon, A., Cattoir, V. and Donnio, P.Y. 2021. Galleria mellonella as a Suitable Model of Bacterial Infection: Past, Present and Future. *Frontiers in Cellular and Infection Microbiology* 11(December), pp. 1–18. doi: 10.3389/fcimb.2021.782733.

Mereghetti, L., Tayoro, J., Watt, S., Lanotte, P., Loulergue, J., Perrotin, D. and Quentin, R. 2002. Genetic relationship between Escherichia coli strains isolated from the intestinal flora and those responsible for infectious diseases among patients hospitalized in intensive care units. *The Journal of hospital infection* 52(1), pp. 43–51. doi: 10.1053/jhin.2002.1259.

Mitsuoka, T. and Hayakawa, K. 1973. [The fecal flora in man. I. Composition of the fecal flora of various age groups]. *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Erste Abteilung Originale. Reihe A: Medizinische Mikrobiologie und Parasitologie* 223(2), pp. 333–342.

Mitsuoka, T., Hayakawa, K. and Kimura, N. 1975. [The fecal flora of man. III. Communication: The composition of Lactobacillus flora of different age groups (author's transl)]. *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Erste Abteilung Originale. Reihe A: Medizinische Mikrobiologie und Parasitologie* 232(4), pp. 499–511.

Mötsch, S. and Schmitt, R. 1984. Replicon fusion mediated by a single-ended derivative of transposon Tn1721. *Molecular and General Genetics MGG* 195(1), pp. 281–287. doi: 10.1007/BF00332760.

Mously, N. Al, Arfaj, O. Al, Fadhil, L. Al and Mukaddam, S. 2016. Antimicrobial susceptibility patterns of ESBL Escherichia coli isolated from community and hospital-acquired urinary tract infections. *Journal of Health Specialties* 4, p. 133.

Mowlds, P. and Kavanagh, K. 2008. Effect of pre-incubation temperature on susceptibility of Galleria mellonella larvae to infection by Candida albicans. *Mycopathologia* 165(1), pp. 5–12. doi: 10.1007/s11046-007-9069-9.

Mukherjee, K., Amsel, D., Kalsy, M., Billion, A., Dobrindt, U. and Vilcinskas, A. 2020. MicroRNAs regulate innate immunity against uropathogenic and commensal-like Escherichia coli infections in

the surrogate insect model Galleria mellonella. *Scientific reports* 10(1), p. 2570. doi: 10.1038/s41598-020-59407-3.

Muldaeva, G., Beisenayeva, A., Arystan, L., Baymanova, A., Haydargaliyeva, L. and Beisenayeva, A. 2019. The Point Prevalence Survey Research of Antibacterial Drugs' Prescription for Outpatient Treatment of Urinary System Infections. *Open access Macedonian journal of medical sciences* 7(15), pp. 2551–2555. doi: 10.3889/oamjms.2019.622.

Muñiz, C.C., Zelaya, T.E.C., Esquivel, G.R. and Fernández Perrino, F.J. 2007. Penicillin and cephalosporin production: A historical perspective. *Revista Latinoamericana de Microbiologia* 49(3–4), pp. 88–98.

Naas, T.; O.S.; B.R.A.; D.M.L.; Z.A.; D.L.; R.P.; I.B.I.,. 2017. *Beta-Lactamase DataBase (BLDB): Structure and Function*. Available at: http://www.bldb.eu/ [Accessed: 24 February 2024].

Nataro, J.P. et al. 1995. Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers. *The Journal of infectious diseases* 171(2), pp. 465–468. doi: 10.1093/infdis/171.2.465.

Nataro, J.P. and Kaper, J.B. 1998. Diarrheagenic Escherichia coli *Clinical Microbiology Reviews* 11(1), pp. 142 LP – 201. doi: 10.1128/CMR.11.1.142.

Nicolas-Chanoine, M.-H. et al. 2008. Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *The Journal of antimicrobial chemotherapy* 61(2), pp. 273–281. doi: 10.1093/jac/dkm464.

Nicolas-Chanoine, M.-H., Bertrand, X. and Madec, J.-Y. 2014. Escherichia coli ST131, an intriguing clonal group. *Clinical microbiology reviews* 27(3), pp. 543–574. doi: 10.1128/CMR.00125-13.

Normark, B.H. and Normark, S. 2002. Evolution and spread of antibiotic resistance. *Journal of internal medicine* 252(2), pp. 91–106. doi: 10.1046/j.1365-2796.2002.01026.x.

Novais, A. et al. 2006. Dissemination and persistence of blaCTX-M-9 are linked to class 1 integrons containing CR1 associated with defective transposon derivatives from Tn402 located in early antibiotic resistance plasmids of IncHI2, IncP1-alpha, and IncFI groups. *Antimicrobial agents and chemotherapy* 50(8), pp. 2741–2750. doi: 10.1128/AAC.00274-06.

Nowak, K. et al. 2017. Highly diverse and antimicrobial susceptible Escherichia coli display a naïve bacterial population in fruit bats from the Republic of Congo. *PloS one* 12(7), p. e0178146. doi: 10.1371/journal.pone.0178146.

Olsen, R.J., Watkins, M.E., Cantu, C.C., Beres, S.B. and Musser, J.M. 2011. Virulence of serotype M3 Group A Streptococcus strains in wax worms (Galleria mellonella larvae). *Virulence* 2(2), pp. 111–119. doi: 10.4161/viru.2.2.14338.

Orskov, F., Orskov, I., Evans, D.J.J., Sack, R.B., Sack, D.A. and Wadström, T. 1976. Special Escherichia coli serotypes among enterotoxigenic strains from diarrhoea in adults and children. *Medical microbiology and immunology* 162(2), pp. 73–80. doi: 10.1007/BF02121318.

ørskov, I., ørskov, F., Bettelheim, K.A. and Chandler, M.E. 1975. Two new Escherichia coli o antigens, o162 and o163, and one new h antigen, h56. withdrawal of h antigen h50. *Acta pathologica et microbiologica Scandinavica. Section B, Microbiology* 83(2), pp. 121–124. doi: 10.1111/j.1699-0463.1975.tb00080.x.

Pallecchi, L. et al. 2007. Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure. *Antimicrobial agents and chemotherapy* 51(4), pp. 1179–1184. doi: 10.1128/AAC.01101-06.

Paul, D., Babenko, D. and Toleman, M.A. 2020. Human carriage of cefotaxime-resistant Escherichia coli in North-East India: an analysis of STs and associated resistance mechanisms. *The Journal of antimicrobial chemotherapy* 75(1), pp. 72–76. doi: 10.1093/jac/dkz416.

Penders, J. et al. 2006. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 118(2), pp. 511–521. doi: 10.1542/peds.2005-2824.

Petersen, A.M., Nielsen, E.M., Litrup, E., Brynskov, J., Mirsepasi, H. and Krogfelt, K.A. 2009. A phylogenetic group of Escherichia coli associated with active left-sided inflammatory bowel disease. *BMC microbiology* 9, p. 171. doi: 10.1186/1471-2180-9-171.

Petronio, G.P. et al. 2020. In vitro and in vivo biological activity of berberine chloride against uropathogenic E. Coli strains using galleria mellonella as a host model. *Molecules* 25(21). doi: 10.3390/molecules25215010.

Petty, N.K. et al. 2014. Global dissemination of a multidrug resistant Escherichia coli clone. *Proceedings of the National Academy of Sciences of the United States of America* 111(15), pp. 5694–5699. Available at: https://doi.org/10.1073/pnas.1322678111.

Piro, A., Tagarelli, A., Tagarelli, G., Lagonia, P. and Quattrone, A. 2008. Paul Ehrlich: the Nobel Prize in physiology or medicine 1908. *International reviews of immunology* 27(1–2), pp. 1–17. doi: 10.1080/08830180701848995.

Pitout, J.D.D. and Laupland, K.B. 2008. Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern. *The Lancet Infectious Diseases* 8(3), pp. 159–166. doi: https://doi.org/10.1016/S1473-3099(08)70041-0.

Pitout, J.D.D., Peirano, G., Chen, L., DeVinney, R. and Matsumura, Y. 2022. Escherichia coli ST1193: Following in the Footsteps of E. coli ST131. *Antimicrobial agents and chemotherapy* 66(7), p. e0051122. doi: 10.1128/aac.00511-22.

Poirel, L., Lartigue, M.-F., Decousser, J.-W. and Nordmann, P. 2005. ISEcp1B-mediated transposition of blaCTX-M in Escherichia coli. *Antimicrobial agents and chemotherapy* 49(1), pp. 447–450. doi: 10.1128/AAC.49.1.447-450.2005.

Prasastha Ram, V. et al. 2022. Antimicrobial Efficacy of Green Synthesized Nanosilver with Entrapped Cinnamaldehyde against Multi-Drug-Resistant Enteroaggregative Escherichia coli in Galleria mellonella. *Pharmaceutics* 14(9). doi: 10.3390/pharmaceutics14091924.

Queenan, A.M. and Bush, K. 2007. Carbapenemases: the versatile beta-lactamases. *Clinical microbiology reviews* 20(3), pp. 440–58, table of contents. doi: 10.1128/CMR.00001-07.

Ramarao, N., Nielsen-Leroux, C. and Lereclus, D. 2012. The insect Galleria mellonella as a powerful infection model to investigate bacterial pathogenesis. *Journal of visualized experiments : JoVE* (70), pp. 1–7. doi: 10.3791/4392.

Ramos, C.P. et al. 2019. Identification and Characterization of Escherichia coli, Salmonella Spp., Clostridium perfringens, and C. difficile Isolates from Reptiles in Brazil. *BioMed research international* 2019, p. 9530732. doi: 10.1155/2019/9530732.

Rastegar Lari, A., Gold, F., Borderon, J.C., Laugier, J. and Lafont, J.P. 1990. Implantation and in vivo antagonistic effects of antibiotic-susceptible Escherichia coli strains administered to premature newborns. *Biology of the neonate* 58(2), pp. 73–78. doi: 10.1159/000243235.

Riccobono, E. et al. 2012. Carriage of antibiotic-resistant Escherichia coli among healthy children and home-raised chickens: a household study in a resource-limited setting. *Microbial drug resistance (Larchmont, N.Y.)* 18(1), pp. 83–87. doi: 10.1089/mdr.2011.0003.

Robicsek, A. et al. 2006. Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nature Medicine* 12(1), pp. 83–88. doi: 10.1038/nm1347.

Rodríguez, I. et al. 2014. Chromosomal location of blaCTX-M genes in clinical isolates of Escherichia coli from Germany, The Netherlands and the UK. *International journal of antimicrobial agents* 43(6), pp. 553–557. doi: 10.1016/j.ijantimicag.2014.02.019.

Roer, L. et al. 2017. Development of a Web Tool for Escherichia coli Subtyping Based on fimH Alleles. *Journal of clinical microbiology* 55(8), pp. 2538–2543. doi: 10.1128/JCM.00737-17.

Rossolini, G.M., D'Andrea, M.M. and Mugnaioli, C. 2008. The spread of CTX-M-type extended-spectrum β-lactamases. *Clinical Microbiology and Infection* 14, pp. 33–41. doi: https://doi.org/10.1111/j.1469-0691.2007.01867.x.

Rowe-Magnus, D.A., Guérout, A.M. and Mazel, D. 1999. Super-integrons. *Research in microbiology* 150(9–10), pp. 641–651. doi: 10.1016/s0923-2508(99)00127-8.

Sahl, J.W., Matalka, M.N. and Rasko, D.A. 2012. Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments. *Applied and environmental microbiology* 78(14), pp. 4884–4892. doi: 10.1128/AEM.00929-12.

Salyers, A.A., Shoemaker, N.B., Stevens, A.M. and Li, L.Y. 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiological Reviews* 59(4), pp. 579–590. doi: 10.1128/mr.59.4.579-590.1995.

Scalfaro, C., Iacobino, A., Nardis, C. and Franciosa, G. 2017. Galleria mellonella as an in vivo model for assessing the protective activity of probiotics against gastrointestinal bacterial pathogens. *FEMS microbiology letters* 364(7). doi: 10.1093/femsle/fnx064.

Schamberger, G.P., Phillips, R.L., Jacobs, J.L. and Diez-Gonzalez, F. 2004. Reduction of Escherichia coli O157:H7 populations in cattle by addition of colicin E7-producing E. coli to feed. *Applied and environmental microbiology* 70(10), pp. 6053–6060. doi: 10.1128/AEM.70.10.6053-6060.2004.

Schlackow, I., Stoesser, N., Walker, A.S., Crook, D.W., Peto, T.E.A. and Wyllie, D.H. 2012. Increasing incidence of Escherichia coli bacteraemia is driven by an increase in antibiotic-resistant isolates: electronic database study in Oxfordshire 1999-2011. *The Journal of antimicrobial chemotherapy* 67(6), pp. 1514–1524. doi: 10.1093/jac/dks082.

Scully, L.R. and Bidochka, M.J. 2005. Serial passage of the opportunistic pathogen Aspergillus flavus through an insect host yields decreased saprobic capacity. *Canadian journal of microbiology* 51(2), pp. 185–189. doi: 10.1139/w04-124.

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N. and Whittam, T.S. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology* 51(5), pp. 873–884. doi: 10.1128/aem.51.5.873-884.1986.

Selander, R.K. and Levin, B.R. 1980. Genetic diversity and structure in Escherichia coli populations. *Science (New York, N.Y.)* 210(4469), pp. 545–547. doi: 10.1126/science.6999623.

Senior, N.J. et al. 2011. Galleria mellonella as an infection model for Campylobacter jejuni virulence. *Journal of medical microbiology* 60(Pt 5), pp. 661–669. doi: 10.1099/jmm.0.026658-0.

Shaikh, S., Fatima, J., Shakil, S., Rizvi, S.M.D. and Kamal, M.A. 2015. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. *Saudi Journal of Biological Sciences* 22(1), pp. 90–101. doi: 10.1016/j.sjbs.2014.08.002.

Shawa, M. et al. 2021. Novel chromosomal insertions of ISEcp1-bla(CTX-M-15) and diverse antimicrobial resistance genes in Zambian clinical isolates of Enterobacter cloacae and Escherichia coli. *Antimicrobial resistance and infection control* 10(1), p. 79. doi: 10.1186/s13756-021-00941-8.

Sherley, M., Gordon, D.M. and Collignon, P.J. 2003. Species differences in plasmid carriage in the Enterobacteriaceae. *Plasmid* 49(1), pp. 79–85. doi: https://doi.org/10.1016/S0147-619X(02)00107-5.

Shulman, S.T., Friedmann, H.C. and Sims, R.H. 2007. Theodor Escherich: The First Pediatric Infectious Diseases Physician? *Clinical Infectious Diseases* 45(8), pp. 1025–1029. doi: 10.1086/521946.

Sir Alexander Fleming - Facts. [no date]. Available at: https://www.nobelprize.org/prizes/medicine/1945/fleming/facts/ [Accessed: 24 February 2024].

SLANETZ, L.W. and BARTLEY, C.H. 1957. Numbers of enterococci in water, sewage, and feces determined by the membrane filter technique with an improved medium. *Journal of bacteriology* 74(5), pp. 591–595. doi: 10.1128/jb.74.5.591-595.1957.

Sloot, R. et al. 2022. Rising rates of hospital-onset Klebsiella spp. and Pseudomonas aeruginosa bacteraemia in NHS acute trusts in England: a review of national surveillance data, August 2020-February 2021. *The Journal of hospital infection* 119, pp. 175–181. doi: 10.1016/j.jhin.2021.08.027.

Smith, J.M., Smith, N.H., O'Rourke, M. and Spratt, B.G. 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* 90(10), pp. 4384–4388. doi: 10.1073/pnas.90.10.4384.

Song, W. et al. 2011. Chromosome-encoded AmpC and CTX-M extended-spectrum β-lactamases in clinical isolates of Proteus mirabilis from Korea. *Antimicrobial agents and chemotherapy* 55(4), pp. 1414–1419. doi: 10.1128/AAC.01835-09.

Stamm, W.E. and Norrby, S.R. 2001. Urinary tract infections: disease panorama and challenges. *The Journal of infectious diseases* 183 Suppl, pp. S1-4. doi: 10.1086/318850.

Stoesser, N. et al. 2016. Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. *mBio* 7(2), p. e02162. doi: 10.1128/mBio.02162-15.

Stoppe, N. de C., Silva, J.S., Carlos, C., Sato, M.I.Z., Saraiva, A.M., Ottoboni, L.M.M. and Torres, T.T. 2017. Worldwide phylogenetic group patterns of Escherichia coli from commensal human and wastewater treatment plant isolates. *Frontiers in Microbiology* 8(DEC). doi: 10.3389/fmicb.2017.02512.

Sturge, G. 2023. *Migration statistics Summary 1 Understanding migration statistics 2 Flows to and from the UK 3 Migrants living in the UK 4 Flows to and from EU countries 5 Migrants living in EU countries 6 Migration between the UK and EU countries 7 Students in migration statistics 8 Appendix*.

Sweeney, N.J. et al. 1996. The Escherichia coli K-12 gntP gene allows E. coli F-18 to occupy a distinct nutritional niche in the streptomycin-treated mouse large intestine. *Infection and Immunity* 64(9), pp. 3497 LP – 3503.

Syed, S.A., Abrams, G.D. and Freter, R. 1970. Efficiency of various intestinal bacteria in assuming normal functions of enteric flora after association with germ-free mice. *Infection and immunity* 2(4), pp. 376–386. doi: 10.1128/iai.2.4.376-386.1970.

Taszłow, P., Vertyporokh, L. and Wojda, I. 2017. Humoral immune response of Galleria mellonella after repeated infection with Bacillus thuringiensis. *Journal of invertebrate pathology* 149, pp. 87–96. doi: 10.1016/j.jip.2017.08.008.

Terawaki, Y., Takayasu, H. and Akiba, T. 1967. Thermosensitive replication of a kanamycin resistance factor. *Journal of bacteriology* 94(3), pp. 687–690. doi: 10.1128/jb.94.3.687-690.1967.

Terry, L.M., Barker, C.R., Day, M.R., Greig, D.R., Dallman, T.J. and Jenkins, C. 2018. Antimicrobial resistance profiles of Shigella dysenteriae isolated from travellers returning to the UK, 2004-2017. *Journal of medical microbiology* 67(8), pp. 1022–1030. doi: 10.1099/jmm.0.000779.

Toleman, M.A. 2017. The Future of Peritoneal Dialysis in a Moving Landscape of Bacterial Resistance. *Peritoneal Dialysis International* 37(2), pp. 134–140. Available at: https://doi.org/10.3747/pdi.2016.00038.

Toleman, M.A. 2018. *Direct in Gel Genomic Detection of Antibiotic Resistance genes in S1 Pulsed Field Electrophoresis Gels*. doi: 10.1007/978-1-4939-7638-6.

Toleman MA, Beidenbach D, J.R. et al. 2002. Molecular characterisation of b-lactamases from Escherichia coli isolated from a multi-centre trial in India. In: *The Forty-second Interscience Conference on Antimicrobial Agents and Chemotherapy*. San Diego, CA: American Society for Microbiology, Washington, DC, USA., p. 114.

Toleman, M.A., Bennett, P.M. and Walsh, T.R. 2006. ISCR elements: novel gene-capturing systems of the 21st century? *Microbiology and molecular biology reviews : MMBR* 70(2), pp. 296–316. doi: 10.1128/MMBR.00048-05.

Toleman, M.A., Mathias, J., Almusallam, A. and Babenko, D. 2020. Analysis of Escherichia coli phylotypes and known sepsis-causing sequence types in UK sewage reveals a direct link between sepsis rates and carriage of pathogenic sequence types in the community. In: *ECCMID 2020*. European Congress of Clinical Microbiology & Infectious diseases, p. 5544.

Toleman, M.A., Spencer, J., Jones, L. and Walsh, T.R. 2012. blaNDM-1 is a chimera likely constructed in Acinetobacter baumannii. *Antimicrobial agents and chemotherapy* 56(5), pp. 2773–2776. doi: 10.1128/AAC.06297-11.

Toleman, M.A. and Walsh, T.R. 2011. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS microbiology reviews* 35(5), pp. 912–935. doi: 10.1111/j.1574-6976.2011.00294.x.

Torres, M.P., Entwistle, F. and Coote, P.J. 2016. Effective immunosuppression with dexamethasone phosphate in the Galleria mellonella larva infection model resulting in enhanced virulence of Escherichia coli and Klebsiella pneumoniae. *Medical microbiology and immunology* 205(4), pp. 333–343. doi: 10.1007/s00430-016-0450-5.

Touchon, M. et al. 2020. Phylogenetic background and habitat drive the genetic diversification of Escherichia coli. *PLOS Genetics* 16(6), p. e1008866. Available at: https://doi.org/10.1371/journal.pgen.1008866.

Tsai, C.J.Y., Loh, J.M.S. and Proft, T. 2016a. Galleria mellonella infection models for the study of bacterial diseases and for antimicrobial drug testing. *Virulence* 7(3), pp. 214–229. doi: 10.1080/21505594.2015.1135289.

Tsai, C.J.Y., Loh, J.M.S. and Proft, T. 2016b. Galleria mellonella infection models for the study of bacterial diseases and for antimicrobial drug testing. *Virulence* 7(3), pp. 214–229. doi: 10.1080/21505594.2015.1135289.

Tsakris, A., Ikonomidis, A., Poulou, A., Spanakis, N., Pournaras, S. and Markou, F. 2007. Transmission in the community of clonal Proteus mirabilis carrying VIM-1 metallo-β-lactamase. *Journal of Antimicrobial Chemotherapy* 60(1), pp. 136–139. doi: 10.1093/jac/dkm138.

Ur Rahman, S., Ali, T., Ali, I., Khan, N.A., Han, B. and Gao, J. 2018a. The Growing Genetic and Functional Diversity of Extended Spectrum Beta-Lactamases. *BioMed research international* 2018, p. 9519718. doi: 10.1155/2018/9519718.

Ur Rahman, S., Ali, T., Ali, I., Khan, N.A., Han, B. and Gao, J. 2018b. The Growing Genetic and Functional Diversity of Extended Spectrum Beta-Lactamases. *BioMed Research International* 2018. doi: 10.1155/2018/9519718.

Valent, P. et al. 2016. Paul Ehrlich (1854-1915) and His Contributions to the Foundation and Birth of Translational Medicine. *Journal of Innate Immunity* 8(2), pp. 111–120. Available at: https://dx.doi.org/10.1159/000443526 [Accessed: 24 February 2024].

Vergis, J. et al. 2019. Antimicrobial Efficacy of Indolicidin Against Multi-Drug Resistant Enteroaggregative Escherichia coli in a Galleria mellonella Model. *Frontiers in microbiology* 10, p. 2723. doi: 10.3389/fmicb.2019.02723.

Vergis, J., Malik, S.V.S., Pathak, R., Kumar, M., Kurkure, N. V, Barbuddhe, S.B. and Rawool, D.B. 2021. Exploring Galleria mellonella larval model to evaluate antibacterial efficacy of Cecropin A (1-7)-Melittin against multi-drug resistant enteroaggregative Escherichia coli. *Pathogens and disease* 79(3). doi: 10.1093/femspd/ftab010.

Vertyporokh, L. and Wojda, I. 2017. Expression of the insect metalloproteinase inhibitor IMPI in the fat body of Galleria mellonella exposed to infection with Beauveria bassiana. *Acta biochimica Polonica* 64(2), pp. 273–278. doi: 10.18388/abp.2016_1376.

Vogt, R.L. and Dippold, L. 2005. Escherichia coli O157:H7 outbreak associated with consumption of ground beef, June-July 2002. *Public health reports (Washington, D.C. : 1974)* 120(2), pp. 174–178. doi: 10.1177/003335490512000211.

Vollmerhausen, T.L., Ramos, N.L., Gündoğdu, A., Robinson, W., Brauner, A. and Katouli, M. 2011. Population structure and uropathogenic virulence-associated genes of faecal Escherichia coli from healthy young and elderly adults. *Journal of medical microbiology* 60(Pt 5), pp. 574–581. doi: 10.1099/jmm.0.027037-0.

Waksman S. 1952. *Streptomycin: background, isolation, properties, and utilization*.

Walsh, T.R., Toleman, M.A. and Jones, R.N. 2007. Comment on: Occurrence, prevalence and genetic environment of CTX-M β-lactamases in Enterobacteriaceae from Indian hospitals. *Journal of Antimicrobial Chemotherapy* 59(4), pp. 799–800. doi: 10.1093/jac/dkl532.

Wang, L., Tkhilaishvili, T., Bernal Andres, B., Trampuz, A. and Gonzalez Moreno, M. 2020. Bacteriophage-antibiotic combinations against ciprofloxacin/ceftriaxone-resistant Escherichia coli in vitro and in an experimental Galleria mellonella model. *International journal of antimicrobial agents* 56(6), p. 106200. doi: 10.1016/j.ijantimicag.2020.106200.

White, A.P. et al. 2011. Intergenic sequence comparison of Escherichia coli isolates reveals lifestyle adaptations but not host specificity. *Applied and environmental microbiology* 77(21), pp. 7620–7632. doi: 10.1128/AEM.05909-11.

Whitman, W.B., Coleman, D.C. and Wiebe, W.J. 1998. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* 95(12), pp. 6578–6583. doi: 10.1073/pnas.95.12.6578.

Whittam, T.S., Ochman, H. and Selander, R.K. 1983. Geographic components of linkage disequilibrium in natural populations of Escherichia coli. *Molecular biology and evolution* 1(1), pp. 67–83. doi: 10.1093/oxfordjournals.molbev.a040302.

Williamson, D.A., Mills, G., Johnson, J.R., Porter, S. and Wiles, S. 2014a. In vivo correlates of molecularly inferred virulence among extraintestinal pathogenic Escherichia coli (ExPEC) in the wax moth Galleria mellonella model system. *Virulence* 5(3), pp. 388–393. doi: 10.4161/viru.27912.

Williamson, D.A., Mills, G., Johnson, J.R., Porter, S. and Wiles, S. 2014b. In vivo correlates of molecularly inferred virulence among extraintestinal pathogenic Escherichia coli (ExPEC) in the wax moth Galleria mellonella model system. *Virulence* 5(3), pp. 388–393. doi: 10.4161/viru.27912.

von Wintersdorff, C.J.H. et al. 2016. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer . *Frontiers in Microbiology* 7.

Wojda, I., Cytryńska, M., Zdybicka-Barabas, A. and Kordaczuk, J. 2020. Insect Defense Proteins and Peptides. *Sub-cellular biochemistry* 94, pp. 81–121. doi: 10.1007/978-3-030-41769-7_4.

Woodford, N., Fagan, E.J. and Ellington, M.J. 2006a. Multiplex PCR for rapid detection of genes encoding CTX-M extended-spectrum β-lactamases. *Journal of Antimicrobial Chemotherapy* 57(1), pp. 154–155. Available at: https://doi.org/10.1093/jac/dki412.

Woodford, N., Fagan, E.J. and Ellington, M.J. 2006b. Multiplex PCR for rapid detection of genes encoding CTX-M extended-spectrum β-lactamases [4]. *Journal of Antimicrobial Chemotherapy* 57(1), pp. 154–155. doi: 10.1093/jac/dki412.

Yasir, M. et al. 2018. Molecular characterization, antimicrobial resistance and clinico-bioinformatics approaches to address the problem of extended-spectrum β-lactamase-producing Escherichia coli in western Saudi Arabia. *Scientific Reports* 8(1), p. 14847. doi: 10.1038/s41598-018-33093-8.

Al Yousef, S.A., Younis, S., Farrag, E., Moussa, H.S., Bayoumi, F.S. and Ali, A.M. 2016. Clinical and Laboratory Profile of Urinary Tract Infections Associated with Extended Spectrum β-Lactamase Producing Escherichia coli and Klebsiella pneumoniae. *Annals of clinical and laboratory science* 46(4), pp. 393–400.

Zahra, R., Javeed, S., Malala, B., Babenko, D. and Toleman, M.A. 2018. Analysis of Escherichia coli STs and resistance mechanisms in sewage from Islamabad, Pakistan indicates a difference in E. coli carriage types between South Asia and Europe. *The Journal of antimicrobial chemotherapy* 73(7), pp. 1781–1785. doi: 10.1093/jac/dky109.

Zhussupova, G., Utepova, D., Orazova, G., Zhaldybayeva, S., Skvirskaya, G. and Tossekbayev, K. 2021. Evaluation of Antibiotic Use in Kazakhstan for the Period 2017-2019 Based on WHO Access, Watch and Reserve Classification (AWaRe 2019). *Antibiotics (Basel, Switzerland)* 10(1). doi: 10.3390/antibiotics10010058.

Zwietering, M.H., Jongenburger, I., Rombouts, F.M. and Van't Riet, K. 1990. Modeling of the bacterial growth curve. *Applied and Environmental Microbiology* 56(6), pp. 1875–1881. doi: 10.1128/aem.56.6.1875-1881.1990.

Chapter 9 Appendices

1. S1-PFGE Buffer compositions.

- Cell suspension buffer: 100mM EDTA, pH 8. Mix 10 mL of 1 M Tris, pH8 with 20 mL of 0.5 M EDTA, pH 8. And add water to make a final volume of 100 mL.
- Tris-Borate-EDTA (TBE) buffer 10X: 108 g Tris base, 55 g boric acid and 40 mL 0.5 M EDTA, pH 8 were added to 700 mL of water. The pH was then adjusted to 8 with concentrated HCl and made up to 1000 mL with water and autoclaved before use. 0.5X TBE was prepared by adding 100 mL of 10X TBE to 1900 mL of water.
- 1/10 TE buffer; 1 mM Tris–HCl, 0.1 mM EDTA, pH 8. Prepare 1X TE buffer: To 10 mL of Tris 1 M, pH 8, add 2 mL of 0.5 M EDTA, pH 8 and then dilute to 1000 mL with water. Dilute 5 mL of 1X TE to 50 mL to produce 1/10X TE buffer.
- 1X S1 buffer: 30 mM sodium acetate pH 4.6, 1 mM ZnSo4, 5% glycerol. Prepare 10X S1 buffer by adding 12.3 g of sodium acetate and 0.92 g Zinc acetate to 200 mL of water. Add 250 mL of glycerol and adjust pH to 4.6. Then add water to 500 mL. Dilute 5 mL of 10X S1 buffer to 50 mL to produce 1X S1 buffer. Store 10X buffer aliquoted at −20 °C.
- Denaturing solution: 0.5 M NaOH, 1.5 M NaCl. Add 20 g NaOH and 87.66 g NaCl to 1 L water.
- Neutralizing solution: 0.5 M Tris–HCl, pH 7.5, 1.5 M NaCl. Dissolve 60.5 g Tris base and 87.6 g NaCl in 800 mL of water adjust to pH 7.5 with concentrated HCl.