# Genetic Evaluation of ESBL *E. coli* Urinary Isolates in Otago

804104
ISURI UMAYA HAPUARACHCHI

# Abstract

The incidence of infections with extended spectrum beta-lactamase (ESBL)-producing *E. coli* in New Zealand is increasing. ESBL *E. coli* most commonly cause urinary tract infections and are seen in both community and hospital patients. The reason for the increasing incidence of ESBL *E. coli* infections is unknown. In this study, 66 urinary ESBL *E. coli* isolates from Otago in 2015 were fully genetically characterised to understand the mechanisms of transmission. The ESBL gene, *E. coli* sequence types, plasmid types, and genetic context (e.g. insertion sequences) of ESBL genes were determined by a combination of whole genome and plasmid sequencing. A bioinformatic pipeline was constructed for the hybrid assembly of Illumina short reads and MinION long reads of ESBL-encoding plasmids. Significant diversity of *E. coli* strains, plasmids, and the genetic context of ESBL genes was seen. This suggests multiple introductions of ESBL resistance genes or resistant bacterial strains accounts for the increased incidence of ESBL *E. coli* in this low prevalence area. Future studies should investigate modes of transmission of ESBL *E. coli* and the genes they encode in Otago.

# Preface

This work would not have been possible without the support and guidance of Dr James Ussher (Supervisor), Xochitl Morgan (Co-supervisor), Ambarish Biswas (Co-supervisor), Rachel Hannaway (Research Assistant).

# Contents

# A LIST OF TABLES

# B LIST OF FIGURES

# C LIST OF ABBREVIATIONS

| | |
|---|---|
| BLAST | Basic Local Alignment Search Tool |
| CIP | Ciprofloxacin |
| CLI | Clindamycin |
| CHL | Chloramphenicol |
| CTX | Cefotaxime |
| CTX-M | Cefotaximase |
| CsgG | Curlin sigma S-dependent growth |
| dNTPs | Deoxy nucleotide triphosphate |
| ddNTP | Dideoxynucleotide Triphosphate |
| DHB | District Health Board |
| ESBL | Extended Spectrum β-lactamase |
| ESR | Institute of Environmental Science and Research (New Zealand) |
| FOF | Fosfomycin |
| GEN | Gentamicin |
| IS | Insertion Sequence |
| KAN | Kanamycin |
| LB | Luria Broth |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight |
| MDR | Multidrug-resistance |
| MIC | Minimum Inhibitory Concentration |
| MLST | Multilocus Sequence Typing |
| NCBI | National Centre for Biotechnology Information |
| NEO | Neomycin |
| ONT | Oxford Nanopore Technologies |
| OXA | Oxacillin-hydrolysing |
| PCR | Polymerase Chain Reaction |
| SGS | Second Generation Sequencing |
| SHV | Sulfhydryl Variable |
| SNP | Single nucleotide polymorphisms |
| SPT | Spectinomycin |
| ST | Sequence Type |
| TEM | Temoniera |
| Tn | Transposon |
| TGS | Third Generation Sequencing |
| WHO | World Health Organisation |
| WGS | Whole Genome Sequencing |
| UV | Ultrviolet |
| ZMW | zero-mode waveguides |

# 1 INTRODUCTION

## 1.1 ANTIMICROBIAL RESISTANCE

Antimicrobials are the cornerstone of modern medicine. In addition to treating various infections, they are essential to prevent infection in many surgical procedures, and to manage infectious complications of immunosuppressive treatment plans (e.g. chemotherapy) (1). Resistance of bacteria to successive classes of antibiotics has emerged and proliferated since the advent of their use, causing treatment failure or adverse patient outcomes (2-5). There is also a shortage of newly developed antimicrobials to combat resistant strains (6,7). The rise of antimicrobial resistance represents an unprecedented threat to modern medicine, with one report commissioned by the United Kingdom government estimating that antimicrobial resistance will be the leading cause of mortality globally by the year 2050, causing an estimated 10 million deaths a year compared to the current estimate of 700,000 deaths per year, and costing the world economy $USD100 trillion (6).

A specific matter of importance is the development of resistance to the most commonly prescribed antibiotics, the β-lactams (5,8). These antibiotics form the backbone of many empiric treatment regimens. Their widespread use has exerted a selective pressure, leading to the development and spread of resistance (9). In gram negative bacteria, production of β-lactamase enzymes, which hydrolyse the antibiotic, is the most common mechanism of resistance (9). The spread of resistance has driven the development of successive classes of β-lactams, which in turn select for corresponding β-lactamase genes (9). One such group of β-lactamases are the extended spectrum β-lactamases (ESBLs) which hydrolyse extended spectrum cephalosporins containing an oxyimino side chain (i.e. cefotaxime, ceftazidime, ceftriaxone), in addition to penicillins, and monobactams, but are inhibited by clavulanic acid (9). ESBL-producing *E. coli* also frequently carry multiple other resistance genes, and as a result are resistant to most other classes of antimicrobials (10).

The World Health Organization (WHO) recently identified ESBL *E. coli* as one of nine resistant bacteria of international concern because of their substantial public health impact (1). *E. coli* are the most frequent cause of urinary tract infections and bloodstream infections, and are one of the leading causes of foodborne gastrointestinal infections (1, 11, 12). As empiric therapy is often ineffective for patients with these cephalosporin resistant infections, infection is associated with increased mortality, as well as associated additional costs in hospitalization, antimicrobial therapy, and medical care compared to cephalosporin susceptible infections (1). In addition, the treatment of ESBL-producing *E. coli* infections frequently requires the use of carbapenems, which are more expensive, less accessible, and of exceptionally broad spectrum. The use of carbapenems also applies a selection pressure for the emergence and spread of carbapenemase producing bacteria, which can be resistant to all available antimicrobials (1,9). In a 2014 report, five out of six WHO regions reported >50% resistance of *E. coli* to 3rd generation cephalosporins, with the highest reported resistance rate being 82% (1).

In comparison, the prevalence of resistance in New Zealand is relatively low but is increasing. In 2016, ESBL producing *Enterobacteriaceae* were isolated from clinical specimens of only 11.1 people per 100000 population in New Zealand, however this had more than doubled between 2007 and 2016 (13). Over the same period, there has been a concomitant increase in the proportion of ESBL-producing *Enterobacteriaceae* that are *E. coli* (52.0% in 2007 to 74.1% in 2016) and a decrease in the proportion that were *Klebsiella spp*. (42.0% in 2007 to 22.3% in 2016) (13). Waitemata, Counties Manukau, and Auckland District Health Boards (DHBs) constitute high prevalence areas, and the DHBs of Lakes, Northland, Waikato, Capital & Coast/Hutt and Taranaki have seen significant increases in the incidence of ESBL-producing *Enterobacteriaceae* between 2009 and 2014 (13). The Southern DHB is a low prevalence area in comparison (13). However, in Otago, the number of ESBL *E. coli* isolates per year implicated in urinary tract infections doubled between 2012 (40 isolates/year) and 2015 (82 isolates/year).

## 1.2 GENETIC UNDERPINNING OF ANTIMICROBIAL RESISTANCE

In order to understand the increase in the overall incidence of ESBL *E. coli*, it is important to consider the genetic diversity that underlies the ESBL phenotype. High diversity would indicate multiple introductions, while low diversity would indicate a single introduction or a limited number of introductions with clonal expansion. An understanding of this genetic diversity is essential to identify reservoirs, sources, and modes of transmission of ESBL *E. coli* and thereby implement the most effective control measures. There are several levels of diversity to consider: *E. coli* clonality including sequence type, ESBL genotypic (gene) variants, plasmid type, and the genetic context (e.g. insertion sequences) of the ESBL gene.

Firstly, clones are highly related bacteria with a recent common ancestor. Different clones are defined by single nucleotide polymorphisms in the core genome (14). The clonality of isolates can be identified by molecular typing procedures, such as multilocus sequence typing (MLST) to define sequence types of *E. coli*. There is not a single sequence type, but multiple sequence types responsible for ESBL production in *E. coli* worldwide. However, there are some predominant sequence types, which are highly virulent, highly transmissible, and multidrug resistant, that have an evolutionary advantage over other *E. coli*, which have contributed to the success of ESBLs. For example, ST131 is the most common worldwide as well as in New Zealand (13 -17). Other common sequence types of ESBL *E. coli* include 69, 73, 95, 38, 393, and 405 (2,15).

The emergence of ESBL-producing *E. coli* ST131 is particularly relevant for the treatment of urinary tract infections, one of the most common infections worldwide, due to its increased adherence and biofilm formation ability, and frequent resistance to fluoroquinolones, a commonly used treatment for urinary tract infections (3,4,15,16). Sequence types 69, 73, and 95 are also commonly associated with urinary tract infections (2). Urinary tract infections are also the most common infection caused by ESBL *E. coli* in New Zealand, consistent with the rest of the world (13).

Secondly, there are multiple families of genes that encode an ESBL. There are over 350 ESBL genotypes, which are grouped into TEM, SHV, and CTX-M type enzymes, as well as the less common VEB, Toho-1, Toho-2, and OXA enzyme types (5,9,18,19). TEM and SHV ESBLs arose first by point mutations in TEM and SHV genes encoding narrow spectrum β-lactamases and became widespread in the 1980s (15,18). So far, 196 TEM variants have been identified, many of them (n = 84) displaying an ESBL phenotype. SHV ESBLs evolved from narrow-spectrum SHV-1 or SHV-11 β-lactamases by point mutation; among the existing 143 SHV variants, 43 display an ESBL phenotype (10). Since the turn of the millennium, CTX-M types have emerged and become dominant globally (5,15,18-20). The diverse CTX-M type ESBLs are classified into 6 groups - 1, 2, 9, 25, 8, and 19LUC - by amino acid sequence similarities (5,15,18). Globally, the most widespread enzymes are CTX-M-14 (group 9) and CTX-M-15 (group 1); these have also been the most common in New Zealand (19-22). Different genes are more common in different parts of the world (15,19,22). For example, CTX-M-2 is prevalent in South America but is rarely seen in the United Kingdom (19,22-24). The most dominant ESBL type in a particular region can change over time (19,22). In Japan, between 2007 and 2012, CTX-M-27 overtook CTX-M-14 as the predominant ESBL type (25). In China between 2009 and 2014, CTX-M-55 overtook CTX-M-15 as the second most common genotype behind CTX-M-14 (26,27). Up to date studies on the proportions of CTX-M gene types are therefore important to gauge any trends and emerging epidemics.

Associations between ESBL encoding genes and sequence types have been well documented worldwide. Overall, CTX-M-15 is the most common ESBL subtype globally. In most countries in the world a strong association between ST131 and CTX-M-15 has been observed; being especially common in North America and Europe compared to Asia (14-17,22,28). *E. coli* ST131 is likely responsible for the global success of CTX-M-15 (14-17,22). In 2013 an ESR study of 224 ESBL-producing *E. coli* from New Zealand, 125 (55.8%) were CTX-M group 1, which includes CTX-M-15, and 97 (43.4%) were CTX-M group 9, which includes CTX-M-14; of the ST131 isolates, there was almost an equal split between group 1 and group 9 CTX-M ESBLs (29). In the 2016 New Zealand nationwide study by ESR, 57.9% (62/107) of *H*30 ST131 had a group 1 CTX-M ESBL, and 42.1% (45/107) had a group 9 CTX-M ESBL, and an association between the *H*30Rx ST131 subclone with CTX-M-15 was found (13). Conversely, an association between CTX-M-27

and ST131 has been seen in some Asian and European countries such as Japan, China, and Germany (22,30-37).

Thirdly, bacterial genomes are plastic and are subjected to point mutations, genome rearrangements (i.e. deletions and insertions, etc.), to localized recombination events, and to acquisition and loss of mobile genetic elements (14). Antibiotic resistance genes encoded on mobile genetic elements may spread at two levels: first some mobile genetic elements are able to transfer between bacterial cells (conjugative mobilizable plasmids and conjugative integrative elements), and second some mobile genetic elements are able to transfer between DNA molecules (insertion sequences, transposons, gene cassettes, and integrons) (2,14,38). The mobile genetic elements encoding resistance mechanisms may vary between isolates of the same sequence type (39). The spread of a mobile genetic element may be more important for the increased prevalence of antimicrobial resistance genes than the spread of a particular bacterial clone (39). Moreover, resistance genes and mobile genetic elements are often found clustered together in large multi-resistance encoding regions in plasmids or 'resistance islands' in the chromosome. Consequently, bacteria are able to acquire a combination of resistance genes, allowing co-selection, and as a result, the success of a particular resistance gene is dependent on its genetic context (38). Thus, characterization of the genetic context, including plasmid studies, is informative when considering the epidemiology of resistance dissemination.

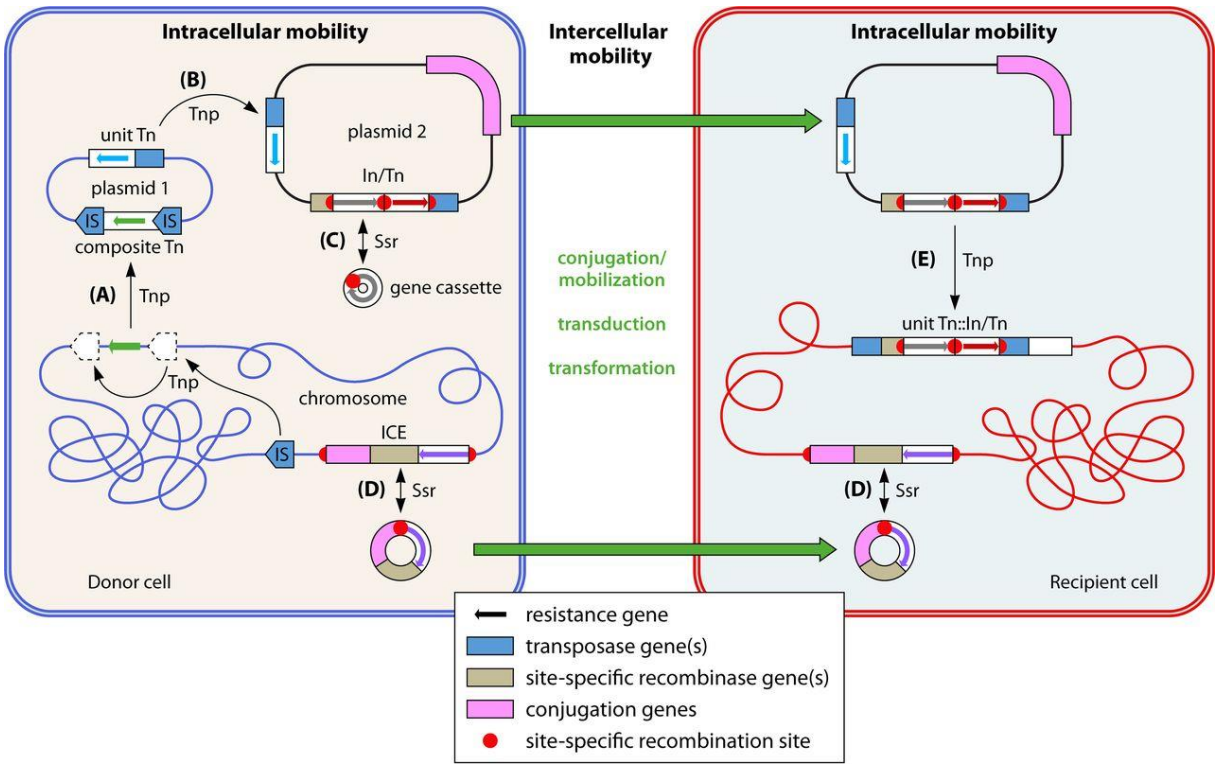



Figure 1: **Examples of mobile genetic elements and processes involved in intracellular mobility or intercellular transfer of antibiotic resistance genes.** Intracellular movement of mobile genetic elements is shown by thin black arrows. Large green arrows indicate horizontal transfer or plasmids or gene cassettes. A: bilateral insertion of insertion sequence elements around a resistance gene allows mobilisation as a composite transposon; B: unit transposon allows mobilization of resistance genes between plasmids or plasmid and chromosome; C: gene cassettes insert into integrons; D: integrative conjugative elements (ICE) capable of intra- or intercellular chromosomal integration; E: transposons and integrons can move into other plasmids or chromosome in the recipient cell. (Figure taken from Partridge et al. 2018 (40).)

ESBL genes are carried by mobile genetic elements, including plasmids, transposons, insertion sequence elements, and integrons. The various mobile genetic elements are illustrated in Figure 1 (40). Plasmids are self-replicating, extrachromosomal circular DNA, some of which can move between different bacterial strains, or even species, by conjugation (41). Plasmids show high degrees of plasticity, carrying several types of mobile elements, such as gene cassettes, integrons, insertion sequence elements, and transposons, which can transfer ESBL genes between plasmids and can also integrate

within the bacterial chromosome (14,38,42). Insertion sequences are the most abundant transposable elements, and the smallest mobile genetic element that carries the genetic information necessary for their own transposition; insertion sequences encode an enzyme, named transposase, which is responsible for mobilization. The transposase recognizes inverted repeat (IR) sequences at the ends of the insertion sequences and is responsible for the integration process (38,42-44). Insertion sequences cause insertion mutations, genome rearrangements and enhance the spread of, and expression of, resistance and virulence genes within species (38,42-44). Unit Transposons are larger than insertion sequences and encode resistance genes as well as transposition genes (Figure 1B) (14,38,42-44). There are three types of transposition: 1) two copies of the same insertion sequence bracketing a mobilized fragment, that form a composite transposon (Figure 1A); 2) mobilization by a single copy of the insertion sequences of adjacent sequences, known as rolling-circle transposition, mediated by insertion sequences including IS*91* and ISCR-like elements; 3) mobilization by a single copy of an insertion sequence of adjacent sequences through a ''one-ended transposition'', where a single insertion sequence recognizes its own inverted repeat left (IRL) and a secondary sequence used as inverted repeat right (IRR) in sequence other than its own, the latter defining the right extremity of the DNA fragment mobilized (44). This third mobilization mechanism has been observed with Tn*21* and IS*Ecp1* (44).

Plasmids, transposons, and insertion sequence elements can carry integrons which in turn can accumulate multiple gene cassettes (41). Gene cassettes consist of a ribosome binding site and *attC* recombination site preceding the gene; they do not encode the mechanism for their own movement, and are the smallest mobile genetic elements associated with antibiotic resistance. While gene cassettes can exist as circular molecules, they are usually inserted into integrons (Figure 1C) (43). Integrons contain an *intI* gene, an *attI* recombination site, and a Pc promoter. The *intI* gene mediates the insertion or release of gene cassettes by encoding an *intI* integrase that catalyses site-specific recombination between the *attI1* site and the *attC* site of the gene cassette. Several cassettes can be integrated into the same integron, and chromosomal integrons can form a reservoir of transferable gene cassettes encoding antimicrobial resistance and virulence factors, resulting in resistance to multiple classes of antimicrobials (41-44).

Plasmid transmission has played a major role in the spread of ESBL resistance genes (4,14,15). Plasmids are extrachromosomal circular DNA, which replicate independently, and some can be transferred between bacteria by conjugation (14,38). Plasmids can be classified by their incompatibility type (45-48). Plasmid incompatibility is the inability of two co-resident plasmids with the same replication and partitioning elements to be stably inherited in the absence of continued selection pressure (46-48). Thus, incompatible plasmids belong to the same incompatibility group (Inc type) (47). Inc type can be determined by simple multiplex polymerase chain reaction (PCR) of specific replicon sequences (45). Several replicons can also exist in one plasmid- called multi-replicon plasmids; this phenomenon is common in IncF groups (47-49). Determining the incompatibility type (Inc type) of plasmids is useful as certain Inc types have been associated with antimicrobial resistance and virulence (45).

ESBL genes are often encoded on plasmids, and CTX-M genes have been found on plasmids of almost all major incompatibility groups (plasmid types) (18). The spread of CTX-M ESBL genes is mainly associated with incompatibility group F, especially the IncFII replicon (19,50,51). IncF plasmids have a narrow host-range, unlike broad host-range replicon plasmids such as IncN, IncI1, and IncL/M that have greater disseminative potential due to their ability to spread the resistance gene across species (14). IncF plasmids are heterogeneous, varying in size (50 to 200 kb), carrying different resistance determinants, and different (combinations of) divergent replicon types. Epidemic resistance plasmids belonging to IncF with divergent replicon types such as FIA, FIB, and FII show high conjugative ability, high transmission rates among *Enterobacteriaceae,* and the tendency to acquire multiple resistance genes, which allows co-selection (4,14,15,19,50-52). The pandemic CTX-M-15 is mainly associated with IncFII. However, there have been recent reports in Turkey, Germany, Egypt, and India, of high incidence of IncFIA and IncFIB plasmids (alone and in combination) (52-57). Broad host range plasmids IncN, IncI1, and IncL/M have also been involved in the dissemination and spread of CTX-M enzymes (19). IncN plasmids conjugate with high frequency, are found in the bacterial flora of humans, animals and the environment (constituting reservoirs of infection), and have been associated with CTX-M-1 (58). SHV-type ESBLs have also been identified on a variety of plasmids, from narrow range (IncF, IncI1, and IncX3), to broad range (IncA/C, IncHI2) (59). These plasmid types have been associated with the

dissemination of SHV-2, SHV-5 and SHV-12. TEM-type ESBLs have been found to be strongly associated with IncA/C plasmids (51,60). The exception to this is the TEM-52 ESBL gene, which has been mainly associated with IncI1 plasmids (44,51,60). IncI1 plasmids are of particular interest as they are highly transmissible, they are globally spread among *E. coli* populations from humans and animals, and they appear to be widely distributed among different members of the *Enterobacteriaceae* family (42). The New Zealand ESR annual surveys do not include plasmid typing, this denotes a gap in the knowledge of ESBL *E. coli* transmission in New Zealand (29).

The mobilization of β-lactam genes has largely been driven by insertion sequences IS*Ecp1*, IS*Cr1*, and IS*26* (2,30,44,61,62). Certain plasmid replicon types such as IncF, IncN, and IncK are associated with certain insertion sequences (e.g., IS*Ecp1* or IS*CR1*), and show greater ability to mobilize CTX-M genes. Insertion sequence elements have been shown to act as strong promoters of CTX-M gene expression (19). In addition, elevated temperature and antibiotic selection pressure have been shown to increase the mobilization frequency of IS*Ecp1* and the ESBL genes present downstream from it; interestingly, cefotaxime, ceftazidime, and piperacillin enhanced transposition, whereas amoxicillin, cefuroxime, and nalidixic acid did not (4,30,63). IS*Ecp1* is the most widely found insertion sequence upstream of different CTX-M genes (associated with all clusters of ESBLs except CTX-M-8), followed by IS*CR1* (associated with members of the CTX-M-2 and CTX-M-9 groups), IS*10*, and less commonly IS*26* (18,30,62,64-67). IS*1* and IS*10*, and even IS*CR1* and IS*26*, have been described interrupting IS*Ecp1* (18,64). Less is known about sequences downstream of CTX-M genes (18,30,44). Downstream of the ESBL genes, different insertion sequence elements have been found in different CTX-M groups. In the CTX-M-1 cluster, sequence orf*477* and mucA have been reported, while in the group 9 cluster IS*903* is common, and insertion sequences are sometimes truncated (Tn*07* and Tn*13*) (64,65,66). The genetic distance between insertion sequences and the CTX-M gene can influence the level of gene expression and hence the level of resistance (18,30). IS*26* and IS*Ecp1* are theorised to be important in *Enterobacteriaceae* multi-resistant plasmid evolution (43,61). IS*26* plays a major role in the mobilization and expression of SHV, being found flanking the gene in composite transposons (43). The *TEM-like* ESBL genes are often carried by three of the earliest bacterial transposons identified. The closely related transposons identified as the carriers of *TEM* genes were

originally designated Tn*A*, and later delineated as Tn*1*, Tn*2*, and Tn*3*. The differences between these transposons are limited to short regions flanking the res site (44). Tn*1* has been associated with *TEM-2*, and Tn*2* has been associated with *TEM-1a* (44).

## 1.3 EPIDEMIOLOGY

While in theory resistance genes associated with particular mobile elements can be acquired from any source, in practice ecology affects the gene flow. The inappropriate use of antibiotics in medicine and agriculture (animal livestock) enforces a selective pressure (68). Antibiotics used in humans and animals can be released into the environment through sewage, wastewater, and agriculture (68). Exposure to antibiotics in the environment selects for mobilization of resistance genes (4,15,43). Antibiotic producing microbes also occur naturally in the environment (68). Plant, soil, wildlife and livestock colonization can constitute reservoirs and sources of antibiotic resistance genes and/or bacteria (68). ESBL producing bacteria have been identified in agricultural (e.g. animal faeces) and industrial waste (human sewage) (68). Biological evidence of clonal linkage can be utilized to support epidemiological evidence (39).

Control strategies include wastewater management (reduced microbial and pharmaceutical pollution), agriculture management (reduction in antibiotic used, treatment of animal waste and reduced access of effluent to water courses), and medical practice management (appropriate antibiotic use, degradable pharmaceuticals). Tracking the spread of infection from person to person is important to define transmission rates, incubation periods, infection duration, antibiotic resistance, reservoirs, sources, and high-risk groups in order to control and predict transmission and the outcome of interventions. Both the WHO and a recent United Kingdom report have stressed the importance of global surveillance (1,6).

In order to implement effective control measures the molecular epidemiology must be assessed to postulate the source and route of transmission (68). Clusters of cases caused by molecularly similar isolates can be analysed to determine the origin (39). A high level of clustering, wherein a few dominant ESBL genotypes, sequence types, plasmid types and common genetic context is observed, indicates a single, or limited number of point source introductions, with clonal expansion and MGE transmission (27,39,69-74). Low level of clustering of isolates, indicates multiple introductions, and may reflect the acquisition of drug resistance through inadequate treatment strategies, rather than transmission of highly virulent drug resistant strains (39,69,70,74).

## 1.4 LABORATORY METHODS FOR CHARACTERISING RESISTANCE

It is indisputable that accurate information regarding the genetic context and diversity of resistance genes are essential to formulating appropriate and effective control measures (1,6). However, this information can be difficult to define in a timely manner for practical reasons as there are many plausible infection pathways and sources (75). Since epidemiologically related pathogens will also be closely related genetically, efficient methods of genetic analysis are required for this purpose (75). Sequencing technology has developed rapidly over the years since the 1960s, from Sanger Sequencing, to Whole Genome Sequencing (WGS), enabling detailed genetic analysis (76).

### 1.4.1 First Generation Sequencing

The first major breakthrough in sequencing technology was Sanger Sequencing, in which chain terminating dideoxynucleotides (ddNTPs) are incorporated into a DNA extension reaction at a concentration such that DNA strands varying by a single nucleotide length for the whole length of the required strand are produced. The ddNTPs are labelled so bases can be differentiated, thus base-calling the whole strand. Initially, ddNTPs were

radiolabelled (four reactions for each base pair and run on a polyacrylamide gel) before fluorometric labelling was introduced (single reaction for all base pairs) (76).

There are major limitations to Sanger Sequencing: it has a long turn-around time, it is labour-intensive, it has a high cost per base pair and a short-read length of <1000 bases (77). This is especially true for whole plasmid or genome resolution which requires a large amount of data for mapping. In a previous study we were unable to completely resolve the ESBL gene types by Sanger Sequencing (ABI BigDye Terminator [version 3.1; Applied Biosystems, Foster, CA, USA] technology), probably due to the read length limitations of Sanger Sequencing (78).

### 1.4.2 **Second Generation Sequencing (SGS)**

SGS, which invariably involves massive parallel sequencing, made efficient whole genome sequencing at a lower cost possible, through modern computational methods. Many methods and technologies have been developed, all with deficiencies in detecting homopolymer sequences. These include ion semiconductor sequencing (Ion Torrent, Life Technologies), as well as the now discontinued pyrosequencing technique (454, Life Sciences) and the Sequencing by Oligonucleotide Ligation and Detection (SOLiD) system (Applied Biosystems). The platform relevant to this study is Illumina Sequencing (San Diego, CA) (76).

The Illumina Sequencing platforms use the principle of bridge amplification on a solid phase substrate (79,80). DNA is enzymatically fragmented and modified by ligation at both ends with adaptors (oligos to attach to the flow cell) and indices (to identify samples for multiplexing). The single stranded DNA library is loaded on to a chip which has synthetic oligonucleotides that function to capture the complementary adaptors on the DNA strands. Next bridge amplification takes place, wherein the DNA strand bends to a bridge shape as the adaptor at the other end of the strand attaches to another complementary oligonucleotide on the flow cell and a polymerase creates a

complementary strand with unlabelled nucleotides, followed by denaturation. This creates distinct clonal clusters (several million of them in each channel) of single-stranded templates anchored to the substrate. Sequencing then begins with the addition of fluorescently labelled reversible terminator deoxy nucleotide triphosphates (dNTPs)of all four bases, primers and DNA polymerase (79,80). Reversible terminator dNTPs are designed so that the fluorophore occupies the 3′ hydroxyl position, which needs to be cleaved away for polymerization to continue. Thus, after the first base incorporation cycle, laser excitation is used to emit fluorescence from each cluster, and a digital picture is taken with a CCD camera. The blocking fluorescent tags are then enzymatically removed and the next complementary reversible terminator dNTP added, continuing the cycle until the sequence of the fragments are determined (76). The natural competition of the four bases minimizes incorporation bias and reduces raw error rate (80).

Illumina Inc. produces a number of devices using this technology. The platform used in this study - recommended for sequencing of microbial genomes - is the Illumina NextSeq 500. This platform has a very high accuracy (0.1% error rate), a sequencing time of 30 hours, a preparation time of 8 hours, and has an output of 100-120 Gb/run output, the cost of platform is $USD250,000, it measures 59 × 53 × 64 cm and weighs 83kg (76,80). Thus, second generation short read sequencing techniques have allowed whole genome sequencing rapidly and affordably compared to first generation sequencing. However, it has not overcome the disadvantage of producing short reads. Averaging at read lengths of 150 bp, SGS has proved deficient in resolving assembly continuity (i.e. positions of elements in a large genomic structure such as a plasmid) due to long repetitive elements, copy number alterations and structural variations. Full portrayal of these structures is important for understanding disease transmission as discussed above (76,81-83).

### 1.4.3 **Third Generation Sequencing (TGS)**

TGS technologies have recently been launched commercially. TGS overcomes the deficiencies in SGS by facilitating sequencing of single molecules, producing long reads and negating the requirement for DNA amplification (76). TGS improves the read length from the scale of tens of base pairs to over 10000 base pairs per read, shortens the turn-around time from days to hours, and circumvents biases introduced by PCR amplification (83).

#### 1.4.3.1 ***Pacbio Sequencing***

Currently, the most commonly used and well validated TGS platform is the single molecule real time (SMRT) platform from Pacific Biosciences, commonly referred to as Pacbio sequencing (76). This technology relies on a chip of arrays of tens of thousands of microfabricated nanostructures known as zero-mode waveguides (ZMWs). Each ZMW is illuminated from below with a wavelength of light that is larger than the aperture of the ZMW, thus penetrating only the lower 20-30nm of the ZMW and acting as a powerful microscope. As phospho-linked nucleotides are incorporated by a DNA template-polymerase complex immobilized at the bottom of the ZMW, the phosphate chain is cleaved, releasing the attached fluorophore and a specific light pulse is produced. The detection volume in the order of nanometres reduces background noise (76,84).

Pacbio offers the advantage over SGS of reasonably accurate long reads in the order of 10 - 40 kb, in a shorter amount of time (4 hours sequencing time), at 0.5-1 Gb per run. However, it is still expensive both upfront ($USD750000 instrument cost) and per output, it is not portable with instrument dimensions of 200 × 77 × 158 cm and 1091 kg weight, and it has a higher error rate than Illumina at 14%. In terms of accuracy, Pacbio has difficulties with resolution of repetitive genomic regions and regions of high GC content, though it can detect modified DNA bases (e.g. DNA methylation patterns) (76).

### 1.4.3.2 *Nanopore Sequencing*

The most recent development in TGS technology is the application of nanopores by Oxford Nanopore Technologies (ONT) in the MinION platform (76,82,85). Nanopore sequencing is a based on the principle of molecules causing characteristic changes in current when passing through a nanopore in a conducting fluid with a voltage potential applied across it. Specifically, electrodes are immersed in chambers on either side of the membrane, setting up a biased voltage across the membrane. An ionic current signal is generated as electrolyte ions pass through the nanopores, which is monitored. When a DNA molecule (negatively charged) passes through a nanopore by drift diffusion, it blocks the pore, and therefore interrupts the current signal. Different base pairs cause characteristic fluctuations in the current (time and amplitude), which can be measured to identify the base pairs (85).

The characteristics of the pore affect the performance as a sequencer. The nanoscopic pore can be solid-state biological transmembrane protein channels embedded in a matrix. Translocation of DNA through the pore must be controlled by coupling an enzyme motor to the nanopore (or another method such as utilizing nucleotide labelling, hairpin termination, or positively charged residues in the nanopore) to allow single nucleotide resolution as natural diffusion is too rapid (>10 nt/μs through solid state pores and >1 nt/μs through biological pores) to differentiate natural current noise from thermodynamic fluctuations (82,86,87). The size of the pore also affects sequencer performance. This is because it determines the number of nucleotides present in the pore at a given time which contribute to the current generated, so the smaller the pore the less noise generated (86). The stability of the pore in different pH and temperature (which may increase during a run) is important for real world use (86).

#### 1.4.3.2.1 Biological Nanopores

Biological nanopores are transmembrane protein channels which are inserted into planar lipid bilayers, liposomes, or other polymer films, and occur commonly in nature

(85,88). They are desirable because pores of precise size and structure (at the atomic level) can be designed using molecular biology techniques (by altering amino acids in existing natural biological membrane sequences) and produced in large numbers. This is not possible with available semi-conductor technology as of yet (85,87).

The first biological nanopore that was proposed for TGS was the alpha haemolysin of *S. aureus,* an exotoxin which is a 232.4 kDa transmembrane channel that can insert itself into planar lipid bilayers, forming a pore measuring 1.4 nm at the narrowest point with a 3.6-nm diameter cap and a 2.6-nm diameter transmembrane β-barrel (85). This makes it close in size to a single stranded DNA molecule. Though this nanopore has advantages in that can it remain functionally stable at temperatures close to 100°C within a wide pH range (pH 2–12), it also has disadvantages in the size being too small for molecules larger than ssDNA and RNA, and too long to distinguish individual nucleotides in single long-chain DNA molecules directly (85).

*Mycobacterium smegmatis* porin A (MspA) is an outermembrane porin which enables hydrophilic nutrient entry in *M. smegmatis*. With a pore measuring ~1.2 nm in diameter and ~0.5 nm in length, it has advantages over alpha haemolysin in that the smaller diameter facilitates trinucleotide sets, its octamer configuration can be used to sequence four nucleotides simultaneously, and it is also more robust to larger pH variations (0-14) at higher temperatures (100$^{o}$C for 30min) (85-87,89).

Bacteriophage phi29 has an advantage over the previous two pore types in that it is wide enough to admit dsDNA, proteins and DNA complexes (3.6 nm-6 nm diameter), has even higher tolerance to pH changes, a wide voltage range, and more options for modifications (85).

Curlin sigma S-dependent growth (CsgG) is the most recent nanopore in commercial use (90). In nature Curli are bacterial surface appendages used for biofilm formation. CsgG is a 262-residue lipoprotein located on the outer membrane that is a prerequisite soluble accessory protein for the secretion of CsgA and CsgB, which form an ungated peptide diffusion channel. CsgG in its channel formation contains a 36 stranded β-barrel domain, forming a transport complex 120 Å (12 nm) in width, 85 Å in height, and a 40 Å inner diameter. There is a restriction formed of three rings of concentric amino acid rings measuring 9 Å in diameter that may guide the extended polypeptide substrate through the secretion pore (90-93).

#### 1.4.3.2.2 Solid State nanopore

Though solid state nanopores are not in use today for long read sequencing, they will probably be the future of nanopore sequencing. This is because of the potential to inexpensively produce high density arrays of nanopores with superior and modifiable chemical, physical and thermal properties, thereby being more robust and durable. They may also be more easily paired with other electronic devices. Focused electron beams are used to make solid state nanopores. However, the technology does not currently exist to fabricate sub-nanometre thick membranes with the control required, with no current leakage through pinholes in ultrathin membranes. Types of materials that show promise include SiN membranes, which have a high chemical resistance and low mechanical stress and are made using an optimized low-pressure chemical vapor deposition, $Al_2O_3$ membranes, with which there is greater control over membrane thickness, and graphene, which has superior mechanical, electrical and thermal properties and natural thickness comparable to nucleotide distance. The main problem however is the fact that they are limited to sensing differences in size alone, instead of chemical properties. This may be overcome by creating hybrid nanopores (85,87).

#### 1.4.3.2.3 The MinION platform

The principle of nanopore biosensors was first conceived in the 1990s (82,85). Proof of concept studies for the use of nanopores coupled with enzyme motors for controlled nucleotide translocation and detection were undertaken in 2005-2010. In 2012 the first platform was announced by Oxford Nanopore Technologies, began beta testing in 2014, and was only released for commercial use in 2015. As is evident, this technology is still in development with improvements being made, methods being finessed, and updates being released in real time (76,82,85).

The platform produced by ONT is known as the MinION, a biological nanopore sequencing USB device, the latest (R9) of which uses an engineered/optimized CsgG derived from *E. coli* (82,90,93). The MinION plugs into a desktop computer or laptop using a high-speed USB 3.0 cable, weighs 90 g, and measures 10 × 3 × 2 cm, making it the first portable sequencing device (83). One flow cell can be used to produce 10-30 Gb of data, and a starter pack including a MinION device, reagents, and flow cell only costs $USD1000. Library preparation time can be as little as ten minutes for some kits. Computational requirements include Windows 7 or above, >8GB RAM and >128GB disk space, and the *MinKNOW* software which carries out data acquisition, real-time analysis, and feedback. The MinION flowcell has 2048 nanopores arranged in 512 channels, allowing up to 512 independent DNA molecules to be sequenced simultaneously. When a run is started wells in each channel are tested, grouped, and ranked in descending order from the most active (G1) to the least (G4). The default run time is 48 hours. G1 is used in the first 24 hours, followed by the other groups (82,83,90,93).

Library preparation is required, though the method complexity and time required differs from kit to kit. Library preparation can involve DNA fragmentation, with methods varying from Covaris g-tube shearing, needle shearing, or enzymatic fragmentation (transposase with barcoded adaptors attached to ends in the same step in the rapid barcoding kit). There is end repair, dA tailing, and hairpin adaptor ligation in some kits. All kits involve the addition of sequencing adaptors and bead purification to remove nucleotides and enzymes. The rapid sequencing kits have the shortest library

preparation time and simplicity, though problems of barcode cross-over between runs using all barcoding kits have been reported (94). There are three sequencing chemistries: 1D, where only one strand is sequenced, 2D (now discontinued), where both strands are sequenced by joining the complementary strands by adaptor ligation, and 1D2, where both strands are sequenced sequentially as the complementary strand is tethered to the membrane (82,83,90,93).

The following are the reasons this new technology is so promising compared to Pacbio and SGS: portable device size, lower price, longer read length, and shorter library prep time. These features can make long read sequencing accessible to many laboratories and can be used in the field. Several studies have already been done where the MinION was used in the field, or methods developed for use in the field. For example, it was used in the detection and characterization of Ebola virus during an outbreak in Liberia, and in the detection of Zika virus and Dengue virus from clinical samples (83,95-97). However, the MinION shows poorer sequence accuracy than Pacbio systems currently. Accuracies ranging from 65%-88% have been reported, though sequencing alignment accuracy has reportedly improved with newer kits and flow cell versions with R9 reaching 97% for 2D/1D$^2$ experiments (83). Though maximum accuracy is still slightly lower for 1D sequencing (94%), 1D sequencing provided much longer sequences than 2D/1D$^2$ (83,94).

## 1.5 WHOLE GENOME SEQUENCE ASSEMBLY AND BIOINFORMATICS

The accuracy of genome assemblies also depends greatly on the downstream bioinformatic analysis methods and tools used. Though the long reads of TGS overcome the problem of low positioning accuracy of SGS, they lack accuracy. Using short read only assembly alone results in more fragmented assemblies. Using TGS reads alone for assembly results in assemblies with lower accuracy, even with polishing steps. The combination of these two methods is therefore beneficial. With hybrid assembly the TGS sequences are used as scaffolds for the more accurate SGS sequences. This results in

higher accuracy and continuity of the assembly, allowing full resolution of the genetic environment (83,98-102).

Hybrid assembly involves the use of several bioinformatic tools in a sequence of steps; forming a 'bioinformatic pipeline'. Firstly, the raw signal generated by the MinION (.fast5 files) must be base called, which produces .fastq files. *Albacore* (Oxford Nanopore Technologies), *Nanocall, Deepnano, Chiron,* and *Deepbinner* are examples of different base callers (103-106). If several samples are run at the same time using biological barcodes, they must be demultiplexed. Other adaptor sequences must also be removed. *Porechop* is an example of a command-line program that is able to do this (107). Some base callers such as *Albacore* (Oxford Nanopore Technologies) and *Deepbinner* also have the ability to demultiplex (106). At this stage reads can also undergo error correction using a program such as *Nanopolish* (108). The long read and short read sequences must then be assembled. Several assemblers have already been developed, each with advantages and disadvantages. There are long read only assemblers such as *Miniasm/Minimap*, which is fast (as little as 2 minutes) but has a high error rate, and *Canu* which is more accurate but has a longer turn around time (up to 2 hours) (99,109). There are also assemblers that can be used for short read only assembly such as *SPAdes* (110). *SPAdes* and *BWAMEM* are examples of programs that can be used for hybrid assembly (110,111). *Racon* and *Pilon* are consensus based assembly error correction software optimized for long reads and short reads respectively (112,113).

## 1.6 WGS IN EPIDEMIOLOGICAL STUDIES

SGS and TGS (including hybrid methods) allow for efficient whole genome sequencing (WGS) of bacteria (76). WGS provides an advantage over older methods of detection for epidemiological studies by allowing identification of resistance genes, virulence factors, plasmids, and phylogenetic linkages (including multilocus sequence typing [MLST]) in one experiment, thereby allowing transmission route elucidation up-to host-host scale, especially in combination with epidemiological population information (75,79).

However, it is important to note that there are many deficiencies in using WGS alone for epidemiological studies. WGS may provide insufficient signal in infections where the mutation rates are low compared to generation time (75). Conversely, there is the possibility of false negative clustering in bacteria with high diversity such as *E. coli* (114). Genomic information also does not necessarily translate to gene expression and transcription, making WGS less accurate than phenotyping in predicting antimicrobial susceptibility due to weak promoters, low copy numbers or unknown mutations (encoded but not phenotypically expressed) (77,115).

There are also bioinformatic difficulties in WGS. The analysis is highly dependent on the quality of sequencing, genome assembly and the chosen reference genome (77). Resistance genes absent in the reference database may be missed when comparing with WGS data (77,115). High computational requirements feature into the costs of WGS that may be overlooked next to instrument and running costs (77). Costs for data storage – in the gigabytes for just one sample - should be considered (cloud-based storage may be impeded by bandwidth, transfer connection strength and cost) (77). Data security and confidentiality should also be taken into consideration (77). Another limitation is the lack of standardization, validation, and comparisons, with all the different bioinformatic tools available (77). Currently the 'bioinformatic bottleneck' may be the biggest impedance for WGS (77). Development of user friendly bioinformatic pipelines to process data is essential for the future (77,115).

## 1.7 AIMS

An increasing incidence of ESBL producing *Escherichia coli* infections has been seen in Otago. I hypothesized that this increase in incidence is due to the clonal spread of one (or a small number of) strains or ESBL-encoding mobile genetic elements. Therefore, the aim of this study was to determine the genetic diversity of ESBL-producing *E. coli* causing urinary tract infections in Otago over a 1-year period (2015-2016). The phylogenetic relationships of bacteria were determined by whole genome sequencing.

To identify the genetic context of the ESBL gene, a bioinformatic pipeline was developed to construct complete plasmid sequences by hybrid assembly of short read sequences (Illumina NextSeq) and long read sequences (MinION Oxford Nanopore Technologies). This information will inform future studies to identify sources and reservoirs of ESBL *E. coli* within the Otago region. It will also provide a bioinformatic pipeline to assist with the analysis of plasmids encoding resistance genes in the future.

# 2 METHODS

## 2.1 CLINICAL ISOLATES

ESBL-producing *Escherichia coli* isolated from urine samples from Otago patients by Southern Community Laboratories (SCL) between February 2015 and January 2016 were included in the study. SCL routinely stores ESBL-producing isolates at -80°C when isolated for the first time; duplicate isolates are not routinely cryopreserved by SCL. Identity and resistance profile of the isolates were reconfirmed as follows to ensure no sampling error has occurred during storage procedure.

As part of a previous study, the identity and resistance profile of isolates were reconfirmed to ensure no sampling error had occurred during the storage procedure (78). Species identification was confirmed by MALDI-TOF MS (Biotyper; Bruker Daltonics, Billerica, MA, USA) (78). Antibiotic resistance (ESBL production and multidrug resistance [MDR]) was confirmed using antibiotic disc diffusion tests according to EUCAST guidelines (78,116). The ESBL group (e.g. CTX-M group 1, group 9 or TEM/SHV/OXA) of the isolates had previously been identified by PCR (78,117). Isolates were tested for 4 common MLST types (ST131, ST69, ST95 and ST73) by PCR (2,78). Following confirmation, isolates were cryopreserved in glycerol stocks (15% glycerol in LB broth) at -80°C. To begin, isolates were subcultured from glycerol stocks and grown/maintained on LB agar.

## 2.2 PLASMID STUDIES

### 2.2.1 Conjugation

Plasmids encoding resistance to extended spectrum cephalosporins were isolated by transfer to an extended spectrum cephalosporin susceptible, sodium azide resistant

recipient laboratory *E. coli* strain; *E. coli* J53, a K12 derived strain. An adapted filter mating conjugation method, as illustrated in Figure 2 was used (118,119).

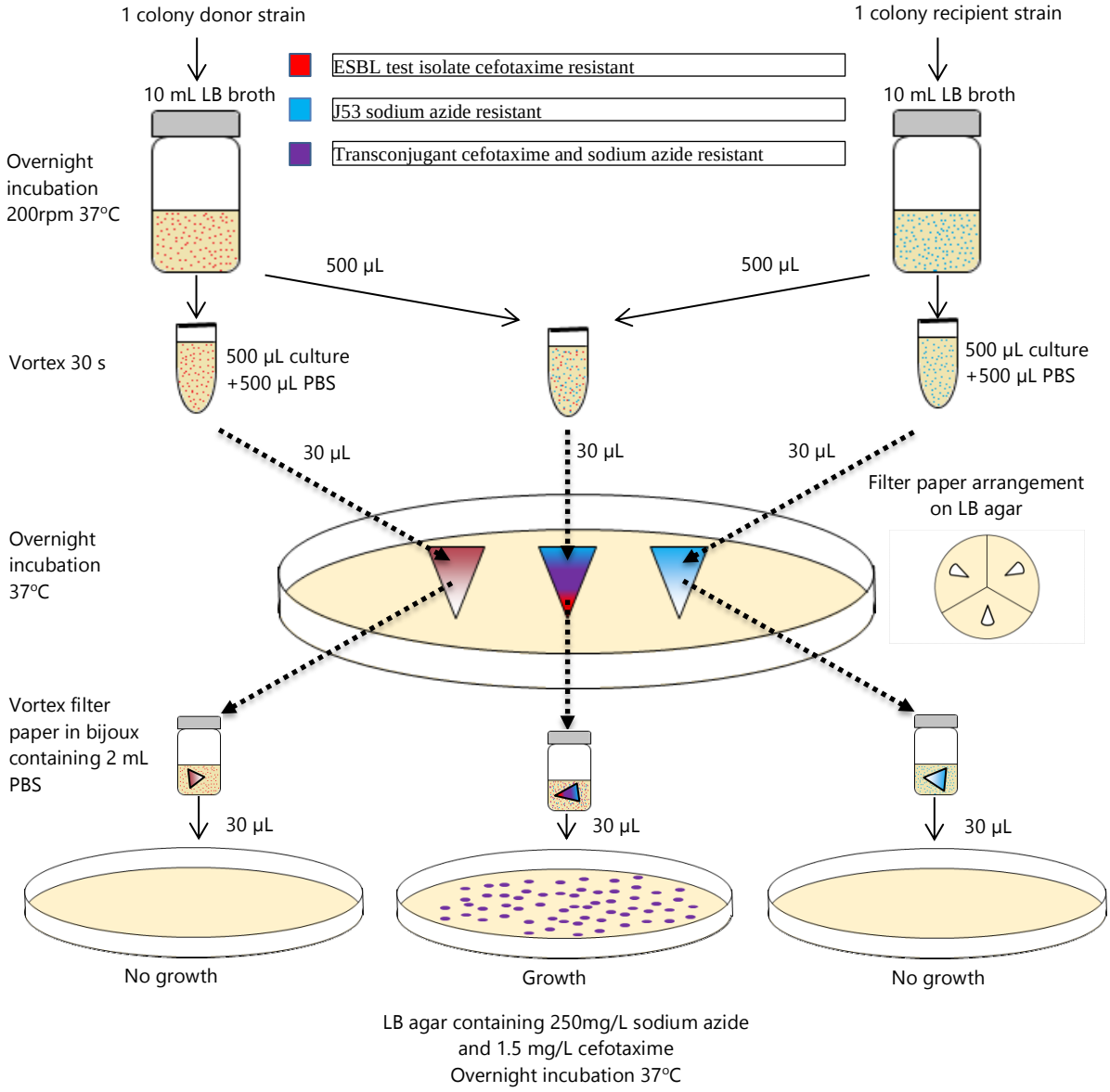



Figure 2: **Filter mating conjugation method**

The optimized method (run on all isolates) was as follows: 1 colony of either the donor (ESBL *E. coli* clinical isolates) or the recipient (sodium azide resistant *E. coli* J53) bacterial strains were inoculated to 10 mL LB broths and were incubated overnight at 37°C , shaking at 200 rpm. The following morning recipient and donor bacteria were mixed in a 1:1 ratio (500 µL: 500 µL) in 1.5 mL Eppendorf tubes, vortexed for 10 s, and 30 µL was pipetted onto a 0.45 µm pore MF-Millipore mixed cellulose ester Membrane

Filter (Merck KGaA, Darmstadt, Germany) placed on plain LB agar as shown in Figure 2. For controls, 30 µL of *E. coli* J53 only or of the test isolate only were also pipetted onto membrane filters. Following overnight incubation at 37°C, filter papers with growth were vortexed in a sterile bijoux bottle containing 2 mL phosphate buffered saline (PBS). 30 µL of this solution was spread onto a LB agar plate containing sodium azide 250 mg/L and CTX 1.5 mg/L and incubated overnight (37°C). Resulting colonies were considered putative transconjugants, were purity plated, and tested by PCR for the ESBL gene that was present in the donor strain and for the IS*5* sequence insertion in the orf*264* gene that is only present in K12 derived laboratory strains (117,120). Glycerol stocks (15% glycerol in LB broth) of the successful conjugants were made and stored at -80°C.

The methodology was refined during the project. In some experiments the ratio of donor to recipient was varied from 1:1 to 1:9. Membrane filters with pore sizes of 0.22 µm and 0.45 µm were compared in some experiments.

#### 2.2.1.1 *Agar dilution method to determine optimal concentrations of sodium azide and cefotaxime*

The ability of *E. coli* J53 and clinical isolates to grow at different concentrations of sodium azide and cefotaxime was determined by the agar dilution method (121). Sodium azide concentrations of 150 mg/L to 400 mg/L and cefotaxime concentrations of 1.5 to 2 mg/L were assessed. All bacterial strains were suspended in PBS to an $OD_{600}$ of 1.0-1.2, equivalent to the absorbance observed for the isolates following the conjugation procedure, and 0.85 µL of this suspension was plated onto agar plates. Plates were incubated at 37°C overnight and then inspected for visible growth.

### 2.2.1.2 *Derivation of sodium azide resistance*

To derive a strain of *E. coli* J53 with greater resistance to sodium azide, the strain was grown to confluence on LB agar supplemented with 150 mg/L sodium azide. A dense suspension was then made in PBS solution from a sweep of the plate. 50 µL of this suspension was then plated on to LB agar containing 200 mg/L or 300 mg/L of sodium azide (122). Colonies from the highest concentration sodium azide-supplemented plate yielding growth were regrown overnight on LB agar containing the same concentration of sodium azide. The above process was then repeated with increasing concentrations of sodium azide (in 25 mg/L increments) until growth was obtained on plates containing 400 mg/L sodium azide.

## 2.3 PLASMID ISOLATION: MODIFIED ALKALINE LYSIS METHOD

Plasmids were isolated from donor strains using a modified alkaline lysis method described by Herringa et al. which is suitable for large plasmids (123). Briefly, strains were grown in 10 mL LB broth overnight, 1.5 mL of which was centrifuged for 10 minutes at 4000xg to pellet the bacteria, then resuspended in 100 µL resuspension buffer (50 mM glucose/10 mM EDTA/10 mM Tris-Cl, pH 8.0) by vortexing. 200 µL of lysis solution (0.2 M NaOH/1% sodium dodecyl sulfate [SDS]) was then added, mixed by inversion and incubated for 5 minutes at room temperature. To this 150 µL 7.5 M ammonium acetate and 150 µL chloroform was added, mixed by inversion and chilled on ice for 10 minutes and then centrifuged at 14,000xg for 10 minutes. The resulting supernatant was transferred to a new microfuge tube with 200 µL precipitation solution (30% polyethylene glycol 8000/1.5 M NaCl), mixed by inversion, chilled on ice 15 minutes, then centrifuged at 14,000xg for 10 minutes to pellet DNA. The supernatant was removed and the DNA pellet resuspended in 50 µL UltraPure™ DEPC-treated water (Thermo Fisher Scientific) at 4$^{o}$C overnight (123). The quality of plasmid DNA obtained was assessed by Nanodrop (Nanodrop one, Thermo Fisher Scientific, Waltham, Massachusetts, USA). OD 260/280 of around 1.8 and OD 260/230 of 2.0-2.2 were considered benchmarks for purity (124,125).

## 2.4 PREPARATION OF ELECTROCOMPETENT BACTERIA

The ESBL carrying plasmids for any isolates that were unable to be conjugated using the filter mating method, were obtained using transformation. Initially electrocompetent *E. coli* ST18 cells were used but showed a low rate of successful transformation (3/6). Electrocompetent *E. coli* DH10B cells (Thermo Fisher Scientific, Waltham, Massachusetts, USA) may be a better host for large plasmids than *E. coli* ST18 so were used for the majority of experiments (126,127).The DH10B electrocompetent strain was prepared in-house as depicted in Figure 3 using the following protocol. SOB broth was prepared with 950 mL $dH_2O$, 20 g BD Bacto™ Tryptone (Becton, Dickenson & Co., NJ, USA), 5 g BD Bacto™ Yeast Extract (Becton, Dickenson & Co., NJ, USA), 0.5 g NaCl, and 10 mL of 250 mM KCl. Two 500 mL flasks of SOB broth were inoculated with 0.5 mL of an overnight culture and grown for 3 hours (or until an OD550 of 0.7 was reached) at 37°C 200 rpm. Cells were then pelleted by centrifugation for 10 minutes at 7000 rpm, washed twice with an equal volume (to the original culture volume) of ice cold 10% glycerol, then once with half volume, then with 30 mL; Cells were pelleted by centrifugation at 7000 rpm for 5 minutes between each wash step. After the final wash step the supernatant was removed and the cells were resuspended in 2 mL of 10% glycerol per litre of original culture. To finish, 40 µL aliquots were transferred into sterile 1.5 mL Eppendorf tubes and frozen at -80°C until use (127).

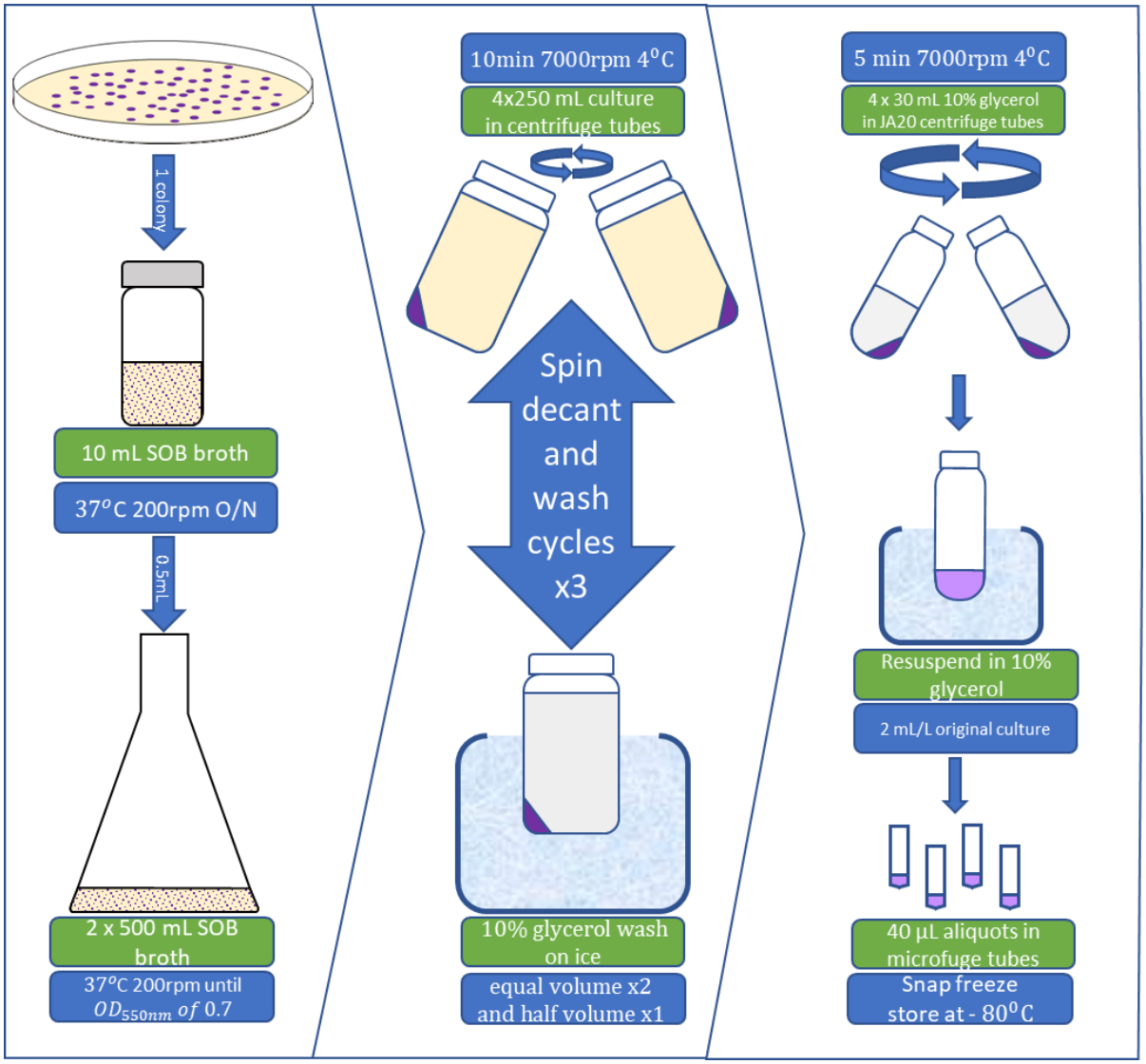



Figure 3: **Preparation of Electrocompetent DH10B cells**

## 2.5 ELECTROPORATION

One microliter of plasmid DNA, containing 40-80 ng DNA unless otherwise specified, was added to a chilled 1.5 mL Eppendorf tube. A 40 µL aliquot of electrocompetent DH10B cells was defrosted for a few minutes on ice and then 20 µL of DH10B cells were added to the Eppendorf containing DNA. Cells and DNA were then transferred to a 0.1 cm electroporation cuvette, and electroporated at 1.8 kV, 200 ohm, 25 µF (standard *E. coli* bacterial settings) using the Gene Pulser Xcell Electroporation system (Bio-Rad Laboratories, Inc, Hercules, CA, USA) (128). Following electroporation, 960 µL of pre-

warmed LB was added to the cuvette containing electroporated cells, then transferred to sterile bijoux, which was incubated for 1 hr at 37°C shaking at 200rpm. Following incubation, 5 µL, 50 µL, and 500 µL were plated on to LB agar plates supplemented with cefotaxime (1 mg/L). Following overnight incubation at 37°C, single colonies were selected, and purity plates made. A negative control (DH10B only) was included in each electroporation run to ensure the DH10B cells did not grow on the LB agar supplemented with cefotaxime. A positive control plasmid (pUC19) was also run (grown on non-selective LB agar) to check the electrocompetence of the DH10B strain. The electrocompetent strain was also streaked on non-selective LB agar to check viability. The presence of the ESBL gene in putative transformants was confirmed by PCR as outlined above, with appropriate negative (DH10B and NTC) and positive controls (clinical isolate). Glycerol stocks (15% glycerol in LB broth) of the successful transformants were made and stored at -80°C. The electroporation method is summarised in Figure 4.

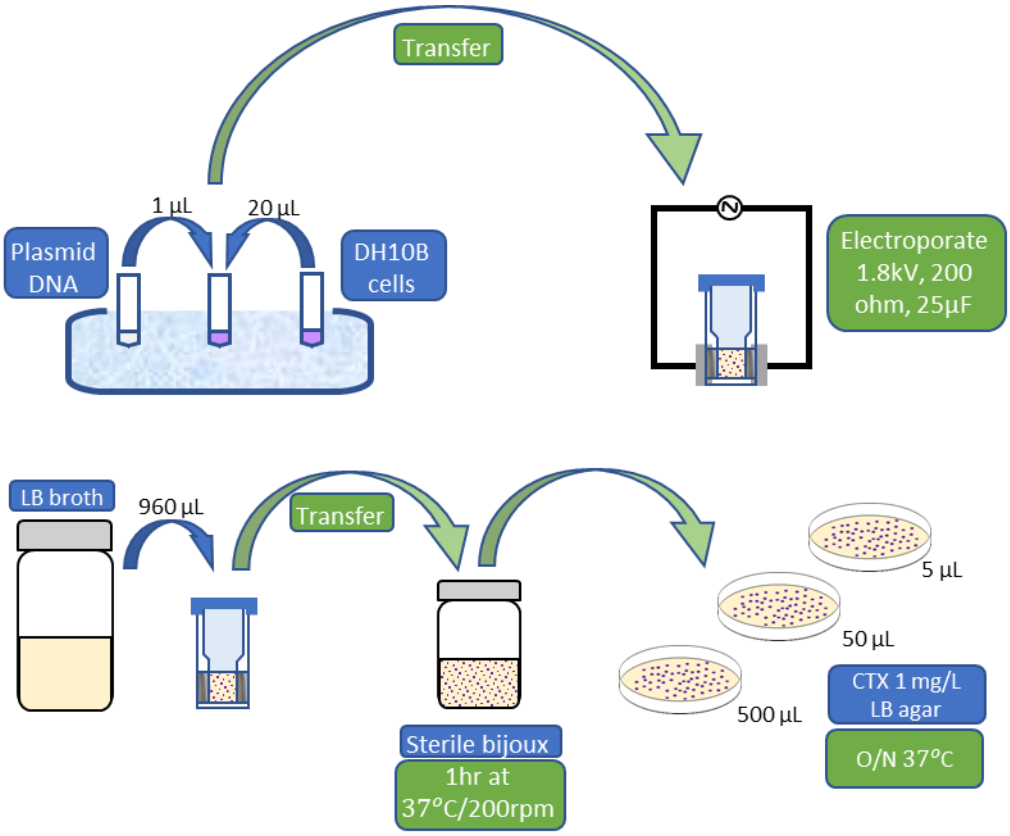



Figure 4: **Electroporation**

## 2.6 CONFIRMATION OF SUCCESSFUL CONJUGATION OR TRANSFORMATION BY PCR

Rapid DNA extraction was performed using the boiling method (117). Briefly, one colony was suspended in 100 µL of PCR-grade distilled water, heated at 95°C for 10 min, then centrifuged at 13,000xg for 10min and the supernatant collected.

For ESBL gene presence confirmation, multiplex PCR assays were performed, and products identified by band size following gel electrophoresis according to procedure and primers described by Dallene et al (117). "Multiplex II" primer mastermix contained forward and reverse primers for amplification/detection of CTX-M groups 1, 2, and 9, "Multiplex I" for TEM, SHV and OXA-like, and "Singleplex 8/25" for CTX-M groups 8 and 25 were used; the primers, product sizes, and references are provided in Table 1 (117). A positive control was used for the CTX-M group 9 primers (CTX-M-15 encoding ST131 *E. coli* from ESR) but was not available for other targets. A no template control was included with each PCR run. The Robust Kapa2G Hotstart PCR kit was used (Kapa Biosystems, Wilmington, MA). Primer mastermix stocks were prepared at final concentrations of 0.2 µM each, except for CTXMGp2_FOR which was prepared at 0.1 µM as per the method of Dallene et al, using TE Buffer (Ambion, Austin, TX) (117).

Table 1: Primers for identifying ESBL genes

| PCR name | β-Lactamase targeted | Primer name | Sequence (5' -3' ) | Primer length (bp) | Annealing Position on target gene | Amplicon size (bp) | Primer concentration (pmol/mL) |
|---|---|---|---|---|---|---|---|
| Multiplex I (TEM, SHV and OXA-1-like) | TEM variants including TEM-1 and TEM-2 | MultiTSO-T_for | CATTTCCGTGTCG CCCTTATTC | 22 | 13–34 | 800 | 0.2 |
| | | MultiTSO-T_rev | CGTTCATCCATAG TTGCCTGAC | 22 | 812–791 | | 0.2 |
| | SHV variants including SHV-1 | MultiTSO-S_for | AGCCGCTTGAGC AAATTAAAC | 21 | 71–91 | 713 | 0.2 |
| | | MultiTSO-S_rev | ATCCCGCAGATA AATCACCAC | 21 | 783–763 | | 0.2 |
| | OXA-1, OXA-4 and OXA-30 | MultiTSO-O_for | GGCACCAGATTC AACTTTCAAG | 22 | 201–222 | 564 | 0.2 |
| | | MultiTSO-O_rev | GACCCCAAGTTTC CTGTAAGTG | 22 | 764–743 | | 0.2 |
| Multiplex II (CTX-M group 1, group 2 and group 9) | variants of CTX-M group 1 including CTX-M-1, CTX-M-3 and CTX-M-15 | MultiCTXMGp1_for | TTAGGAARTGTG CCGCTGYA | 20 | 61–80 | 688 | 0.2 |
| | | MultiCTXMGp1-2_rev | CGATATCGTTGG TGGTRCCAT | 21 | 748–728 | | 0.2 |
| | variants of CTX-M group 2 including CTX-M-2 | MultiCTXMGp2_for | CGTTAACGGCAC GATGAC | 18 | 345–362 | 404 | 0.1 |
| | | MultiCTXMGp1-2_rev | CGATATCGTTGG TGGTRCCAT | 21 | 748–728 | | 0.2 |
| | variants of CTX-M group 9 including CTX-M-9 and CTX-M-14 | MultiCTXMGp9_for | TCAAGCCTGCCGA TCTGGT | 19 | 299–317 | 561 | 0.2 |
| | | MultiCTXMGp9_rev | TGATTCTCGCCGC TGAAG | 18 | 859–842 | | 0.2 |
| Singleplex I (CTX-M group 8/25) | CTX-M-8, CTX-M-25, CTX-M-26 and CTX-M-39 to CTX-M-41 | CTX-MG8/25_for | AACRCRCAGACG CTCTAC | 18 | 172–189 | 326 | 0.2 |
| | | CTX-MG8/25_rev | TCGAGCCGGAASG TGTYAT | 19 | 497–479 | | 0.2 |

*Table 1 taken from Dallenne et al. (117).

The mastermix for PCR reactions was prepared as outlined in Table 2. Each reaction included 1 µL of DNA template or, for the no template control (NTC), 1 µL of PCR-grade water (Invitrogen, NY, USA). ST131/CTX-M group 9 control strain was used as a positive control for Multiplex II. Amplification was carried out according to the KAPA2G Robust Hotstart operating procedure as follows: initial denaturation at 95°C for 3 min; 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1 min; and a final elongation step at 72°C for 1 min.

Table 2: KAPA2G Robust Hotstart kit mastermix

| **Component** | **Volume per 25 µL reaction** |
|---|---|
| KAPA2G Buffer B | 5 µL |
| dNTP Mix 10mM | 0.5 µL |
| Primer Mastermix [a] | 18.4 µL |
| 5 U/µL KAPA2G Robust Hotstart DNA polymerase | 0. 1 µL |

[a] Final primer concentrations as per Table 1

PCR products were run on 2% agarose gel (Molecular grade agarose gel [Thermo Fisher Scientific, Waltham, Massachusetts, USA] and Bionic buffer [Sigma-Aldrich, St. Louis, MO, USA]), supplemented with 10 µL (to 100 mL agar) SYBR safe (Thermo Fisher Scientific, Waltham, Massachusetts, USA), at 100 V for 1 hour. A 100-1000 bp ladder: GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used as products ranged from 100 bp-900 bp. The gel was visualized with an ultraviolet (UV) gel imager.

PCR was also used to confirm that the putative transconjugants were the J53 strain of *E. coli*, which is derived from the laboratory K12 strain (120). The method described by Kuhnert et al targets the IS*5* sequence insertion in the orf*264* gene, which is present in K12 strains, as illustrated in Figure 5 (120). The primers and their targets are shown in Table 3 (120). PCR for K12 was performed using the KAPA2G Robust Hotstart kit, with the mastermix as above (Table 2). The following conditions were used: denaturation at

95°C for 3 min; 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1 min; and a final elongation step at 72°C for 1 min. The sodium azide-resistant strain *E. coli* J53 strain was used as a positive control and one of the clinical strains, and the group 9 *E. coli* from ESR were used as negative controls.

Gel electrophoresis was performed with 1% agar (Seaplaque low melting point agarose: Lonza, Rockland, ME, USA + Bionic buffer Sigma-Aldrich, St. Louis, MO, USA), supplemented with 10 µL SYBR safe (Thermo Fisher Scientific, Waltham, Massachusetts, USA), and 1 Kbp plus DNA Ladder (Invitrogen Life Technologies, Carlsbad, CA, USA). The gel was visualized with a UV gel imager.

Table 3: K12 identification primers

| Primer | Sequence | Positions | Reference Gene |
|---|---|---|---|
| K12-R | 5'-ATCCTGCGCACC AATCAACAA-3` | 508 to 488 | orf264 |
| K12-L | 5'-TTCCCACGGACATGAAGACTACA-3` | 21 to 43 | |
| K12IS-L | 5`-CGCGATGGAAGATGCTCTGTA-3` | 293 to 313 | IS5 |

*Table 3 taken from Kuhnert et al. (121)

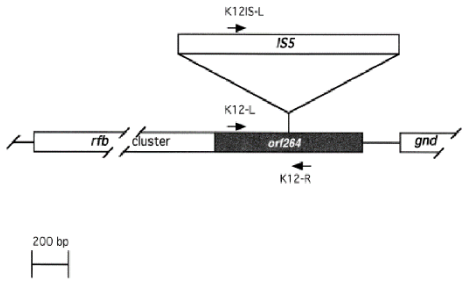



Figure 5: **K12 PCR primer locations.** Arrows show locations of primers used for K12 PCR, and boxes represent genes. orf*264* (shaded box), which encodes rhamnose transferase, harbours an IS*5* sequence in most K-12 strains specifying the rfb-50 mutation. The K12-L primer is dependent on the complete gene orf*264*. K12IS-L primer is dependent on the 3' end of orf*264* and the presence of IS*5*; K12-R is the reverse primer for both K12-L and K12IS-L forward primers. *Figure 5 taken from Kuhnert et al. (121)

## 2.7 DNA SEQUENCING WITH MINION

Plasmids from conjugants and transformants where extracted and sequenced using the MinION (Oxford Nanopore Technologies Ltd, Oxford, UK).

### 2.7.1 Preparing Input DNA

Plasmids of the conjugates and transformants were isolated using the modified alkaline lysis method described above (123). Plasmid quantity and quality was assessed using the Nanodrop one (ThermoFisher Scientific, MA, USA). OD 260/280 of around 1.8 and OD 260/230 of 2.0-2.2 were considered benchmarks for purity. Plasmids were then cleaned using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) and 80% ethanol. Briefly, 54 µL (1.8x) of AMPure XP beads at room temperature were mixed with 30 µL plasmid DNA in a 1.5 mL Eppendorf tube, incubated for 5 minutes at room temperature, placed in magnetic rack (Figure 6), and allowed to clear (as beads are drawn to the magnet) (129). The supernatant was removed, then 200 µL of freshly prepared 80% ethanol was added to tube, incubated for 30 seconds, then removed; this wash step was repeated. The beads were air dried, removed from magnet, then resuspended in 40 µL of DEPC water and incubated for 5 minutes at room temperature to elute DNA. The tube was then placed on magnet again and clear supernatant containing DNA was removed to a clean Eppendorf tube.


Figure 6: **Magnetic Separation Rack.** Figure taken from reference (129).

One microliter of cleaned plasmid DNA was diluted 1 in 10 with DEPC water to measure quantity and quality. The quality was assessed with 1 µL on the Nanodrop. The Qubit Fluorometer (Invitrogen) was used to accurately determine the DNA quantity following the manufacturer's protocol. Briefly, 1 µL of diluted DNA was mixed with 199 µL working solution (prepared by adding 1 µL Qubit HS dsDNA reagent to 199 µL Qubit dsDNA HS buffer) and incubated for 3 minutes at room temperature. The two standards provided (prepared by adding 10 µL standard to 190 µL working solution, incubated for 3 minutes at room temperature) were used to calibrate the Qubit before measurement. The remaining 8 µL of cleaned diluted DNA was run on 0.8% low melting point agarose gel at 70 V for 15-50 minutes to check for unwanted shearing.

### 2.7.2 **DNA Library Preparation**

The DNA library was prepared using the Rapid Barcoding Sequencing kit (SQK-RBK004) (Oxford Nanopore Technologies Ltd, Oxford, UK) according to the protocol provided by the manufacturer. This kit comes with 12 barcodes; that is 12 specimens can be run per batch. Kit components (stored at -20$^{o}$C) were thawed at room temperature and centrifuged briefly prior to use.

DNA was diluted to 400 ng/7.5 µL in DEPC water in a volume of (at least) 10 µL. The concentration was checked using the Qubit Fluorometer (Invitrogen) using the procedure described above. Then 7.5 µL of this diluted DNA (containing 400 ng DNA) was combined with 2.5 µL of one of 12 specific barcoded fragmentation mixes (RB01-RB12: one for each sample) in a 0.2 mL thin walled PCR tube and mixed by gently flicking the tube. The mixtures were incubated at 30$^{o}$C for 1 minute then at 80$^{o}$C for 1 minute in a thermocycler, then transferred to ice to cool. The barcoded samples were then pooled at equal (1:1) ratio (10 µL each).

As more than four samples were barcoded at a time, a clean-up procedure was carried out in order to increase throughput as recommended by the ONT protocol. Briefly, an equal volume of AMPure XP beads at room temperature was added to the pooled sample, mixed by flicking the tube, and incubated on a Hula mixer (Invitrogen) for 5 minutes at room temperature. Then the sample was pulse spun and pelleted on a magnet, and the supernatant was removed. The beads were then washed twice with 200 µL 80% ethanol as previously described. The pellet was then resuspended in 10 µL of 10 mM Tris-HCl pH 8.0/50 mM NaCl, incubated for 2 minutes at room temperature, pelleted using the magnet, and the supernatant containing DNA transferred to a clean 1.5 mL Eppendorf DNA LoBind tube. 1 µL of rapid adaptor (RAP) was added to 10 µL of barcoded DNA, mixed by flicking the tube, and incubated at room temperature for 5 minutes. The prepared library was then stored on ice until ready to load onto the MinION flow cell.

DNA library preparation methodology was refined during the study. Samples were run in four batches. One batch of fresh plasmids that were never frozen following extraction was compared to a batch where plasmids were frozen between quality testing and running on the MinION. Batch 4 was run on a new flow cell.

### 2.7.3 **Running the MinION**

MinION sequencing was performed using R9.4 (FLO-MIN106) flow cells (Figure 7) (130). Sequencing was run for 5 – 12 hours. Oxford Nanopore Technologies *MinKNOW* software (versions 0.45.2.6 - 2.34.3) was used to collect raw electronic signal data.

The MinION was plugged in to the computer and the MinION graphic user interface (GUI) turned on. Quality control (QC) was run on the flow cell before each run. Buffer containing a QC DNA molecule is present in the flow cell on arrival (this buffer is also provided separately for storage). The *MinKNOW* software (0.45.2.6 - 2.34.3) recognizes the distinctive signal of the QC DNA molecule, allowing it to validate the integrity of the

nanopore array, and outputs the estimated number of available (good quality) channels in groups of four for the experiment (130).

Sequencing buffer, loading beads, flush buffer, and flush tether reagents (provided in the kit) were thawed at room temperature, then placed on ice. Sequencing buffer and flush buffer were mixed by vortexing, flush tether was mixed by pipetting, then all were briefly centrifuged and returned to ice. Loading beads were mixed by pipette just prior to loading.

The MinION was then opened and the priming port on the flow cell was opened (clockwise). A small volume was drawn back from the port to remove any bubble by turning the wheel of a P1000 pipette (set at 200 μL) to show 220-230 μL, whereupon a small volume of buffer enters the pipette tip. After visually confirming that there was continuous buffer from the priming port across the sensor array, the flow cell priming mix was prepared by mixing 30 μL of thawed and mixed Flush tether with thawed and mixed flush buffer. Then 800 μL of the priming mix was loaded into the flow cell via the priming port and left for 5 minutes.

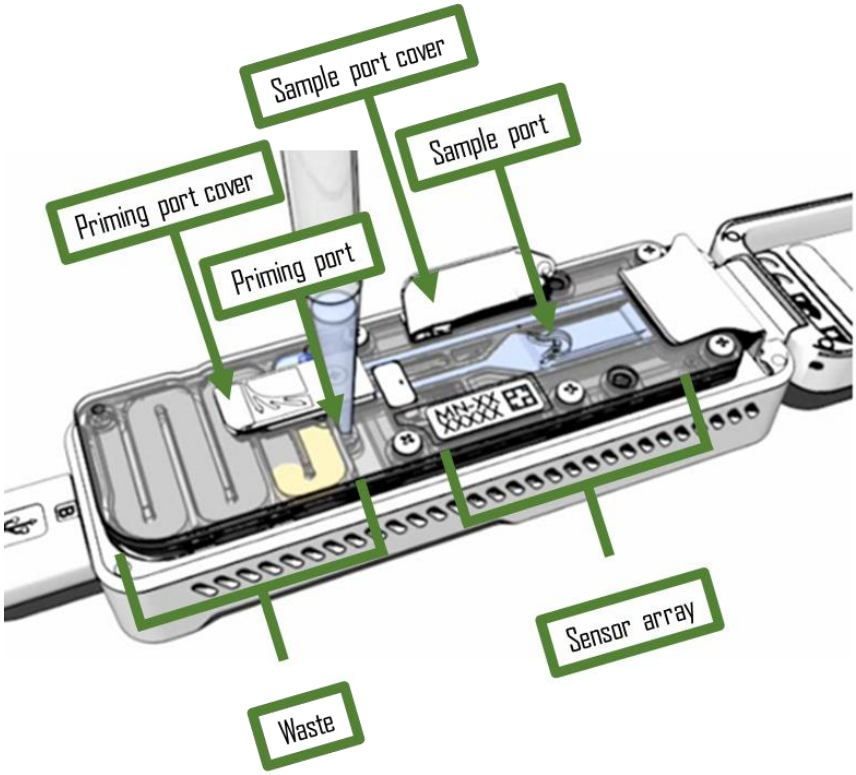



Figure 7: **Parts of the MinION flowcell.** Figure adapted from reference (130)

The sample loading library was then prepared by adding 34 µL Sequencing Buffer (SQB) to an Eppendorf tube, followed by 4.5 µL Nuclease-free water. The Loading Beads (LB) were mixed by pipetting up and down then 25.5 µL was added to the mixture. Finally, 11 µL of the prepared DNA library was added to the mixture. The sample port cover was then opened and a further 200 µL of priming mix was added to the priming port (Figure 7). The sample library was mixed by flicking the tube prior to loading. Then the 75 µL sample library was added to the sample port dropwise without touching the port. Sample and priming ports, and MinION lid were closed, and the sequencing run was begun using the *MinKNOW* software (version 0.45.2.6 - 2.34.3).

During the run, several parameters were checked on the *MinKNOW* software (version 0.45.2.6 - 2.34.3) including the number of active pores (ideally within 10-15% similarity to the active pores called during QC), heatsink temperature (ideally 34°C), the read length histogram, and pore occupancy (ideally 70% of all active pores in the first 30

minutes). The run time was selected to be 5 hours, and live-base calling was used for the first batch. For the two subsequent batches the run time was increased to 12hrs in order to increase output and live-base calling was not used (due to the increased run time to prevent loss of data from any connection issues). The last batch was run for 5 hours but live base calling was not utilized. *EPI2ME* software (version 2.52- 2.58) was used to analyse the MinION data. The flowcell was washed between each use by adding 150 µL of solution A followed by 150 µL of solution B. 500 µL of storage buffer was added before storing the flowcell.

## 2.8 ILLUMINA NEXTSEQ WHOLE GENOME SEQUENCING DATA ANALYSIS

All 66 isolates underwent whole genome sequencing (WGS) on an Illumina NextSeq instrument (San Diego, CA) at the Microbiological Diagnostic Unit Public Health Laboratory (MDU PHL), University of Melbourne. DNA extraction and library preparation were performed at the MDU PHL. The resulting data was analysed in this project. The *Nullarbor* (version 1.20) bioinformatic pipeline was used to determine the sequence type, the resistance genes present, and the phylogenetic relationship of isolates from the WGS data (131). In the *Nullarbor* pipeline, *Trimmomatic* (version 0.27) was used to clean the reads, then de novo assembly was conducted with *SPAdes* (version 3.10.1) (131,132,110). Species identity was detected with *kraken* (version 0.10.5-beta). The sequence type was determined by in silico MLST with *mlst* (2.8), which automatically detects the appropriate scheme (131). The resistance genes were defined with *Abricate* (version 0.4) (131,133). Variants were identified with *Snippy* (version 3.2-dev) (134). *FastTree,* along with *Snippy* (version 3.2-dev) was used to determine the phylogenetic relationship of isolates based on core genome single nucleotide polymorphisms (SNPs) using the *Escherichia coli* ST131 strain EC958 complete genome (Genbank: HG941718.1) as the reference genome (134,135). Lastly, the accessory genome was compared with *Roary* (version 3.11.3) (136). *Nullarbor* outputs a .html format report that summarizes the quality of reads, species identification, MLST, and

resistance profile of each isolate as well as pan genome and core genome phylogeny trees, with core SNP distances and density.

## 2.9 MinION Data Analysis

The following describes the post analysis data processing of the MinION reads, from quality control to basecalling the MinION reads to the hybrid assembly of MinION reads and Illumina reads. Figure 8 summarizes the finished pipeline (137).

### 2.9.1 Basecalling and Demultiplexing

The first batch was base-called in real-time using the *MinKNOW* software (versions 0.45.2.6 - 2.34.3) (Oxford Nanopore Technologies, Oxford, UK), which produces fastq files. Subsequent batches were not live-base-called: output was in the format fast5. *Albacore* (version 2.3.0) (Oxford Nanopore Technologies, Oxford, UK) was used to base-call these fast5 files, converting them to fastq (see Appendix A Fig 1). *Porechop* (version 0.2.3) was used to demultiplex and to remove adaptor sequences; a shell program (See Appendix A Fig 2) was written to partially automate the process (107).

### 2.9.2 Quality Assessment

*EPI2me* (version 2.52- 2.58) (Oxford Nanopore Technologies, Oxford, UK) was used to assess the quality of each sample. *EPI2me* provided mean and median read length, read number, and read quality for each barcode. These parameters were compared between each batch.

### 2.9.3 **Hybrid Assembly (Data Analysis)**

Hybrid assembly using MinION reads as scaffolds and Illumina reads for error correction was used to construct complete plasmid sequences. Methodology was refined during the study. Three program options were trialled and compared using a subset of isolates. First *Miniasm/Minimap* (version 0.2) (a long-read only assembler) followed by *SPAdes* (version 3.11.1) (used as a hybrid read assembler) (see Appendix A Fig 3) (110,138). Second, the '–plasmid' option of *SPAdes* (version 3.11.1) (see Appendix A Fig 3). Third, *bowtie2* (version 2.3.4.1) in combination with the *Unicycler* pipeline (version 0.4.6) (See Appendix A Fig 4-Fig 6) (139,140).

*Bowtie2/Unicycler* was concluded to be the best method and was used in the final version of the hybrid assembly pipeline (139,140). *Bowtie2* (version 2.3.4.1) is a tool that aligns reads to reference genomes. *Unicycler* is an assembler specialized for combining Illlumina and MinION data (139). *Unicycler* (version 0.4.6) uses a specialized algorithm involving the use of *SPAdes* (version 3.11.1), *Miniasm* (version 0.2), *Racon* ( version 1.2.1), BLAST+ (version 2.2.31), *Pilon* (version 1.22), *Java* (version 1.8.0_151), *bowtie2* (version 2.3.4.1), and *samtools* (version 1.4.1) (139). *Unicycler* (version 0.4.6) was used to perform hybrid assembly of all the plasmids.

#### 2.9.3.1 ***bowtie2 filtering***

The whole genome of each isolate was sequenced on the Illumina. However, only the plasmids encoding the ESBL gene that were able to be isolated were sequenced with the MinION. As this difference in data may cause difficulties or errors during assembly, *bowtie2* filtering (version 2.3.4.1) was used to prefilter the Illumina reads to include only those that match the relevant plasmid as sequenced by the MinION (140). Any hybrid assembly that did not yield a complete plasmid with *Unicycler* alone was repeated with *bowtie2* filtering.

As *bowtie2* requires fasta (.fa) files to build reference a database, *Miniasm* (version 0.2) was used to generate a long read only assembly (from the MinION reads) (Appendix A Fig 4) (138). As long read only assembly has a higher error rate, and filtering Illumina reads on this assembly may filter out relevant Illumina sequences, affecting the accuracy of downstream analyses, *sed* was used instead of *Miniasm* to convert the long read fastq files to fasta files of any plasmids that were incomplete with *Miniasm* (Appendix A Fig 4).

The *bowtie2* build function (version 2.3.4.1) was then used to make an index from the resulting output file (from *Miniasm* or *sed*). Then *bowtie2* (version 2.3.4.1) was run to give an output of paired Illumina reads (forward and reverse files) that matched to the index. The '--very-sensitive local' pre-set was used to account for the higher error rate in the MinION reads. The paired output files of *bowtie2* can be used just as unfiltered paired *Illumina* reads as input for *Unicycler*.

### 2.9.4 **Analyses on Completed Plasmids**

Following successful plasmid assembly, the resulting graph (.fastg output) was viewed on *Bandage* (version 0.8.1), which facilitates the manual curation of assemblies based on coverage and connections between contigs, to determine the location and multiplicity of resistance genes and to allow manual annotation (141). Plasmids were typed *in silico* using *PlasmidFinder* (version 1.3-2.0) and *pMLST* (version 1.4) (the full sequence of the circularized plasmids in .fa format were used as input) (142). *ISfinder* (version 2.2.31), was used to identify insertion sequence, which were then manually annotated on the *Bandage* graph (143). NCBI blast (version 2.8.0), *EPI2me* antimicrobial resistance application (version 2.2.14), *ResFinder* (version 3.1), and CARD antimicrobial resistance (version 4.2.2) were used to determine the presence of antibiotic resistance genes (144-148).

A sunburst chart of the MinION data was created to visualize the diversity of the isolates. The pan genome tree *Nullarbor* output was labelled with the MinION data (and the sequence type from the *Nullarbor* output) to visualize the relatedness of the isolates.

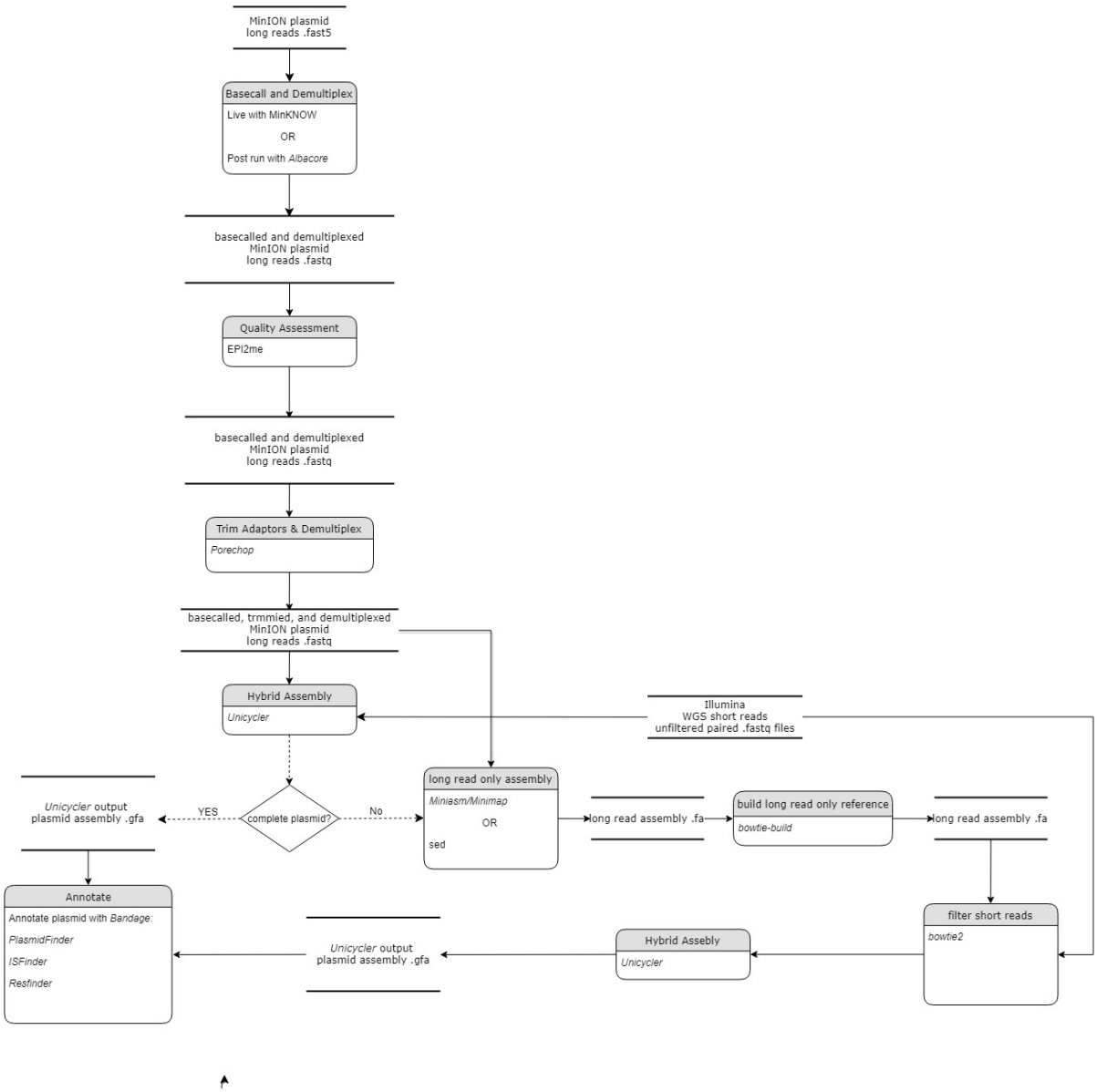



Figure 8: **Bioinformatic pipeline graph**

## 2.10 STATISTICAL ANALYSIS

Results were tabulated and analysed statistically. The $\chi^2$ test and *post hoc* binomial test and sign test was used for statistical analysis. A two-tailed $p < 0.05$ was considered significant. All analyses were performed using the statistical software SPSS (build 1.0.0.1246; SPSS Inc., Chicago, IL, USA) and Excel Analyse-it for Microsoft Excel (version 2.20, Analyse-it Software, Ltd. https://analyse-it.com/; 2009).

# 3 RESULTS

## 3.1 CLINICAL ISOLATES

There were 140 ESBL *E. coli* isolated by SCL from urine samples from Otago (from both hospital and community patients) between February 2015 and January 2016 (inclusive). 58 of these were duplicates, 10 were never stored, and 3 were missing from storage. Of the 69 non-duplicate ESBL *E. coli* urinary isolates collected, 66 were confirmed to be ESBL *E. coli* by MALDI-ToF MS and by susceptibility testing (78). Of the 3 isolates that were not ESBL *E. coli*, 2 were found to be AmpC producers, rather than ESBL-producers, and 1 was cefpodoxime susceptible and not an ESBL (78). This may reflect sampling error during storage in the presence of a double population of ESBL and non-ESBL-producing bacteria. Most isolates (75.7%, n=50) were of community-origin.

## 3.2 WHOLE GENOME SEQUENCING: ILLUMINA RESULTS

Overall sequence quality was flagged as acceptable in all but one isolate. The number of reads ranged from 1,371,106 to 6,021,814, with an average read length of 148-149 bp, and average depth of coverage of 75-172 fold (excluding the low-quality isolate which had a depth of 39 fold). All isolates had a Phred score of 33; a Phred score of greater than 30 has an incorrect base call probability of 1 in 1000 (Supplementary Data, Appendix B).

### 3.2.1 Illumina Species Identification

K-mer analysis against the *kraken* database confirmed the identity of all isolates as *E. coli*, with the first match for all isolates being *E. coli*, at 33.24% to 82.23% identification, and the second match for all isolates being <8.28%. The second match was unclassified in 62/66 (94%) cases, the other four being *S. sonnei*, *K. pneumoniae*, and (two) *S.*

*enterica*. Even though the *Nullarbor* pipeline flagged the quality of reads for use in species identification to be low (<80%) in all but 13 of the isolates, the gap between the first and second match (most of which were 'unclassified) was large in all isolates. The median identity classification percentage for the first match was 72.69%. There were 8 isolates with identity classification percentage of less than 50% (33.24%-46.95%). These all had low quality scores as assessed by *Nullarbor*. There were 14 other isolates with low quality scores with identity classification percentage of 50.85-63.99% (Supplementary Data, Appendix B).

### 3.2.2 **SNP Phylogeny**

Using *E. coli* ST131 (HG941718.1) as the reference genome, 6 -7 clusters were observed in the core genome phylogeny tree (Figure 9) (Supplementary Data, Appendix B). Several smaller clusters (9-10) were evident in the pan genome phylogeny tree (Figure 10) (Supplementary Data, Appendix B). Several clusters which are separate in the pan genome tree appear to be closely related in the core genome tree. For example, the large cluster of 21 isolates (closely related to the reference genome) in the core genome tree, is separated into 3 disparate clusters in the pan-genome tree - one sub-group of 11 isolates (isolate 68, 69, 15, 78, 12, 11, 82, 27, 44, 21, and 20), another sub-group of 2 (isolates 36 and 58) and a further subgroup of 7 (isolates 8, 76, 85, 60, 73, 35, 40). In the other instance, a large cluster in the core genome (consisting of isolates 67, 54, 79, 66, 75, 29, 38, 19, 26, 62, 39, 41, 45, and 72) appears as three separate sub-groups in the pan genome tree: one group of 6 isolates (isolates 67, 54, 79, 66, 75, and 71), another group of two isolates (isolates 29 and 38) and the remainder forming a group of 7. Isolate 10 is an example of where the pan genome tree indicating a single isolate with a differing accessory genome to those in the same cluster in the core genome tree.

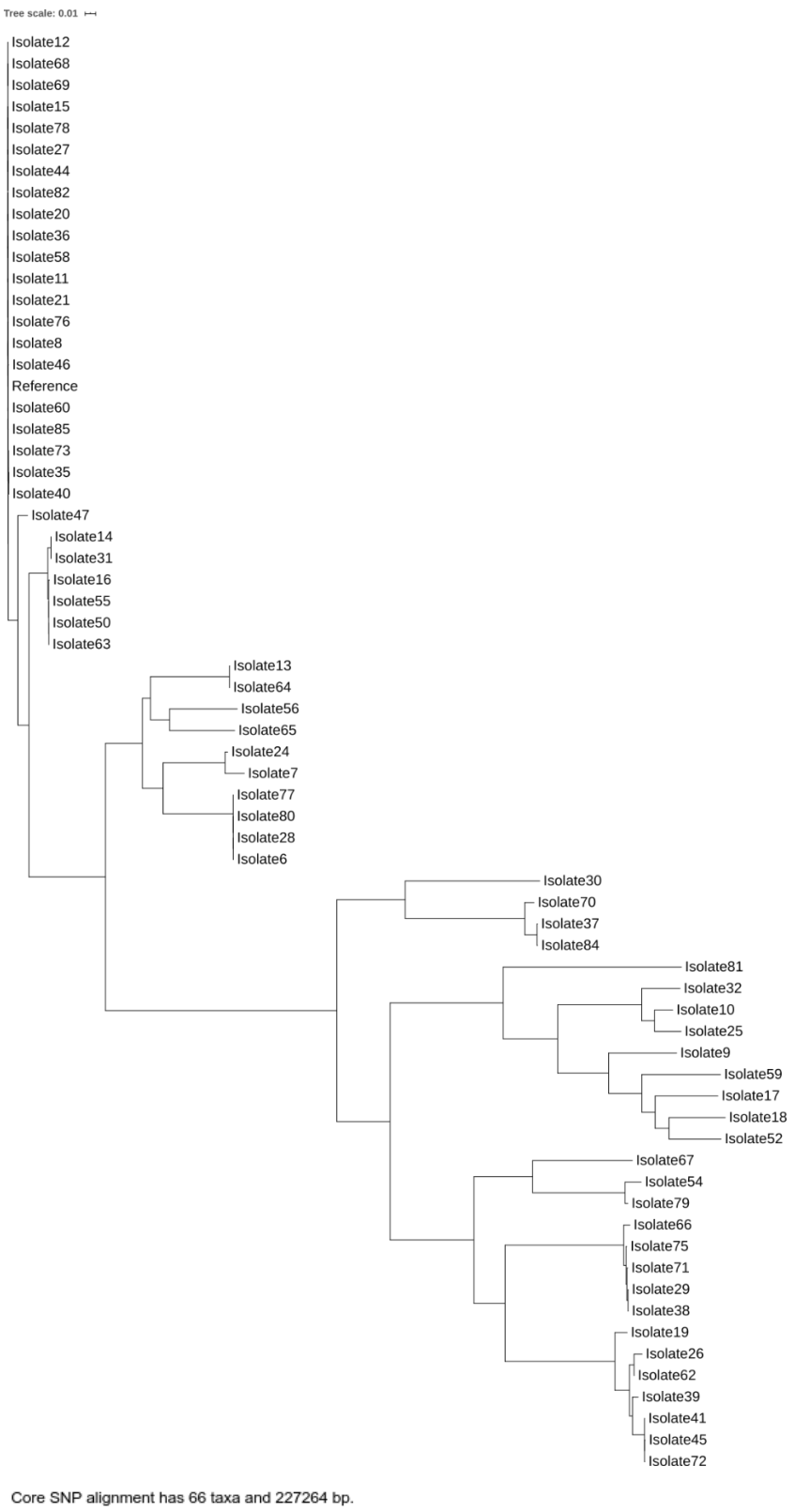



Figure 9: **Core SNP phylogeny.** Produced by *FastTree* (version 2.1.9 Double precision [No SSE3]) with SNP from *Snippy-core* (version 3.2-dev) (in the Nullarbor pipeline) using a reference genome of ST131 ESBL *E. coli* (Genbank: HG941718.1). Graph generated using iTOL online (version 3.0) reference (149)

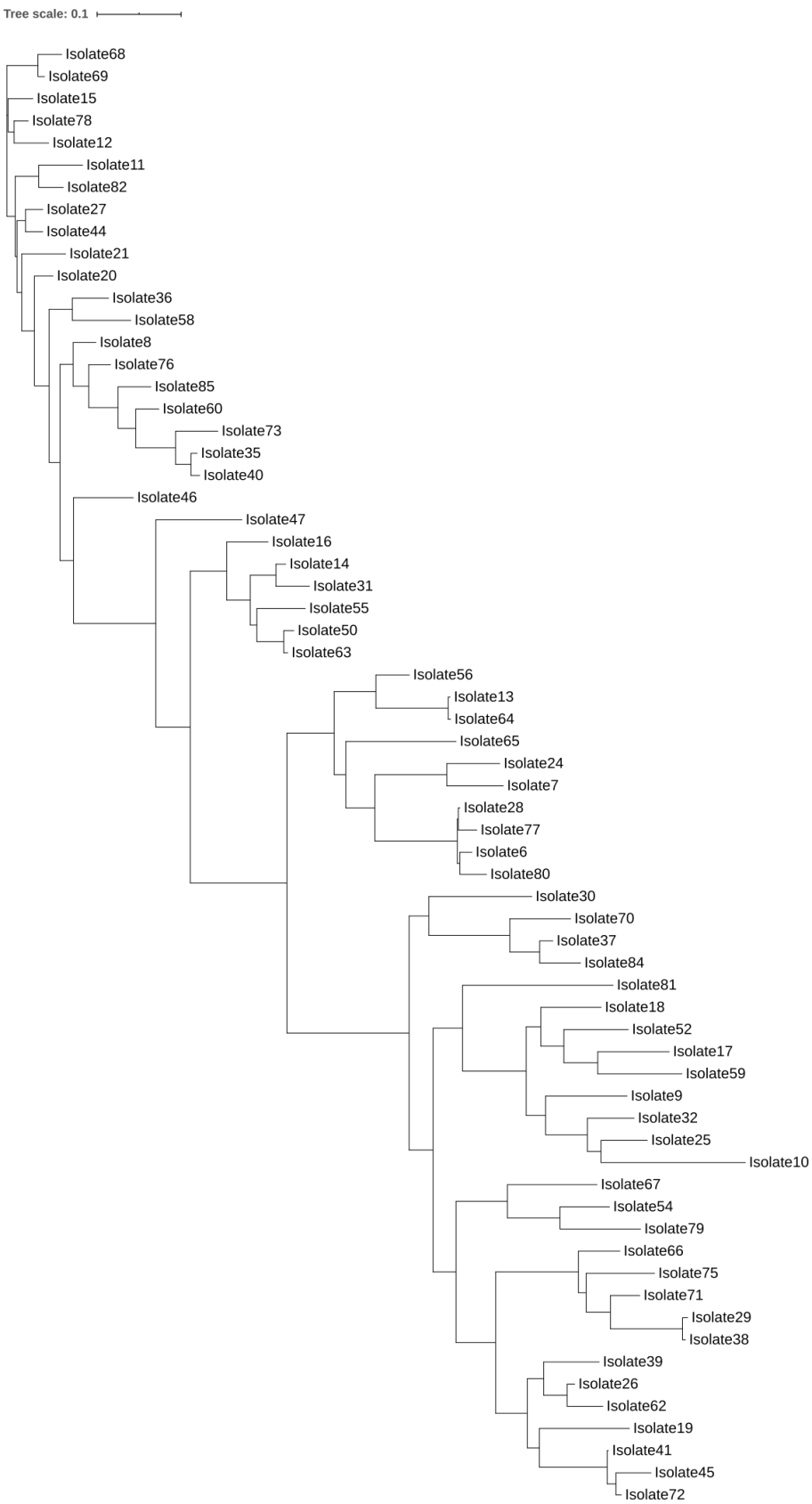



Figure 10: **Pangenome SNP phylogeny.** Produced by *FastTree* (version 2.1.9 Double precision [No SSE3]) and *Roary* (version 3.11.3) (in the Nullarbor pipeline). Graph generated using iTOL online (version 3.0) reference (149)

### 3.2.3 **MLST**

MLST was performed by *mlst* (2.8) using the Achtman *E. coli* scheme, which includes seven conserved housekeeping genes: *adk, fumC, gyrB, icd, mdh, purA* and *recA* (Supplementary Data, Appendix B) (131,150). There were 23 different sequence types (ST) identified. ST131 was the most common ST (27/66, 40.9%) followed by (in descending order) ST38 (7/66, 10.6%), ST405 (5/66, 7.6%), ST636 (4/66, 6.1%), ST648(3/66, 4.5%), ST127 (2/66, 3.0%), and one each of 15 different STs were found (ST10, ST12, ST14, ST69, ST95, ST101, ST162, ST297, ST393, ST410, ST448, ST457, ST744, ST1193, ST3036, and ST4204). The sequence type of 2 (3.0%) isolates was unable to be determined; the *Nullarbor* report had flagged these sequence reads as low quality (Supplementary Data, Appendix B). These isolates had previously been identified as ST69 and ST131 by PCR in a previous study (78). The MLST data is summarized in Figure 11.

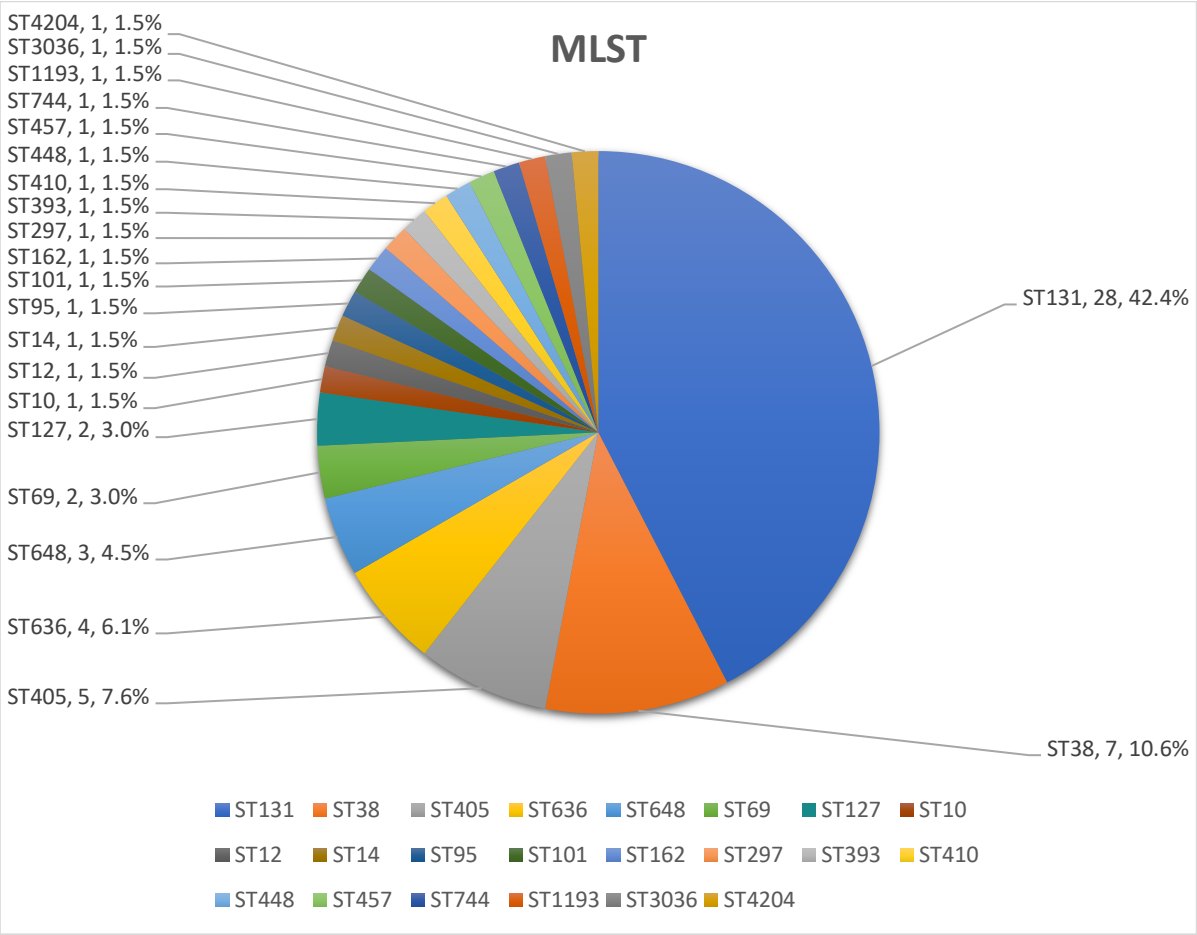



Figure 11: **MLST of ESBL *E. coli* isolates.** This figure includes the ST131 isolate and the ST69 isolate that were previously identified by PCR (78) with the MLST data gleaned from WGS.

## 3.2.4 Resistome

### 3.2.4.1 *ESBL Genes and other β-lactamase genes*

The β-lactamase genes for CTX-M-27, CTX-M-14, CTX-M-15, CTX-M-55, TEM-1B, TEM-1A, SHV-12, OXA-1 and OXA-9, were identified alone and in combination (Table 4) (Supplementary Data, Appendix B). In addition, there was one CMY-2 AmpC resistance gene found in conjunction with CTX-M-15 and TEM-1B. Of note, neither of the OXA variants identified, nor TEM-1 are known to have an ESBL phenotype (144-146). SHV-12 is known to have an ESBL phenotype (144-146).

The majority of isolates (36/66, 54.5%) contained TEM, SHV and/or OXA genes according to WGS. However most (27/36, 75%) were in combination with CTX-M genes (which are likely to be solely responsible for the ESBL phenotype), with only 9 being found alone. Most of the OXA genes were OXA-1 (10/12, 83.3%), with only two OXA-9 (2/12, 16.7%). These two OXA-9 were the only OXA genes found in combination with TEM1B and SHV, and not in combination with a CTX-M gene. It should also be noted that no SHV genes were found in combination with CTX-M genes, with SHV found in most of the isolates not containing CTX-M genes (8/9, 88.9%). Instead a CTX-M/TEM-1B combination was the most common (17/27, 63.0%), with only 6 (22.2%) containing the CTX-M/OXA combination, and a further 4 (14.8%) with a CTX-M/OXA/TEM1B combination. These combinations were found across all CTX-M groups. CTX-M-14 and CTX-M-55 were found only in combination with TEM1B, while both OXA and TEM1B appeared in CTX-M-27 and CTX-M-15 carrying isolates. CTX-M-27 was notable as the only CTX-M gene type where the majority did not occur in combination with TEM1B or OXA genes; 14/18 (77.8%) of CTX-M-27 isolates carried CTX-M-27 only).

There were more group 9 CTX-M ESBL genes (35/66, 53.0%) than group 1 CTX-M ESBL genes (20/66, 30.3%), though this difference was not significant ($p > 0.05$). While the group 9 CTX-M genes were roughly evenly split between CTX-M-14 (17/35, 48.6%) and CTX-M-27 (18/35, 51.4%), the majority of group 1 CTX-M genes were CTX-M-15 (15/20, 75%) with only 7 (35%) CTX-M-55 genes. In two isolates (3.0%) CTX-M-14 genes were found in conjunction with CTX-M-55; these isolates are considered separately for clarity. Though CTX-M-14 genes were the most common CTX-M gene there was no significant difference between the proportions of CTX-M-27, CTX-M-14 and CTX-M-15 ($p > 0.05$).

Table 4: Beta lactamase gene proportions

| Beta-lactamase encoding genes | Number | Percentage (%) |
|---|---|---|
| Total CTX-M | 57 | 86.4 |
| CTX-M-14/55 | 2 | 3.0 |
| Total Group 9 | 35 | 53.0 |
| Total CTX-M-14 | 17 | 25.7 |
| CTX-M-14 | 7 | 10.6 |
| CTX-M-14/TEM-1B | 10 | 15.2 |
| Total CTX-M-27 | 18 | 27.3 |
| CTX-M-27 | 14 | 21.2 |
| CTX-M-27/TEM-1B | 2 | 3.0 |
| CTX-M-27/OXA-1 | 2 | 3.0 |
| Total Group 1 | 20 | 30.3 |
| Total CTX-M-15 | 15 | 21.2 |
| CTX-M-15 | 5 | 7.6 |
| CTX-M-15/OXA-1 | 4 | 6.1 |
| CTX-M-15/OXA-1/TEM-1B | 4 | 6.1 |
| CTX-M-15/TEM-1B | 2 | 3.0 |
| Total CTX-M-55 | 5 | 7.6 |
| CTX-M-55 | 2 | 3.0 |
| CTX-M-55/TEM-1B | 3 | 4.5 |
| Total SHV/TEM/OXA only | 9 | 13.6 |
| TEM1B | 1 | 1.5 |
| SHV-12 | 1 | 1.5 |
| SHV-12/TEM-1B | 5 | 7.6 |
| SHV-12/TEM-1A/OXA-9 | 2 | 3.0 |

### 3.2.4.2 ***Relationships between ESBL genes and MLST***

Figure 12 shows the MLST types of the isolates carrying different ESBL genes. In accordance with the high number of sequence types identified in this study, there was high diversity of MLST types observed carrying each ESBL gene, with all CTX-M gene types being found in isolates of different sequence types. All isolates encoding CTX-M-55

alone were of different sequence types, namely ST38, ST95, ST131, ST1193, and ST4204, though the two CTX-M-55/CTX-M-14 encoding isolates were both ST69. However, associations between MLST and ESBL genes were also observed. While ST131 isolates carried all types of CTX-M genes, ST131 was the most frequent MLST carrying CTX-M-15 (7/15, 46.7%) and CTX-M-14 (6/17, 35.3%). ST131 isolates also represented the majority of isolates carrying CTX-M-27 (72.2%, 13/18), and a statistically significant association of ST131 with CTX-M-27 was observed ($p < 0.01$) by $\chi^2$ test. There were only 3 other MLST types that carried CTX-M-27, compared to 6 and 7 respectively for CTX-M-14 and CTX-M-15. The numbers of other STs and CTX-M types were too small for statistical analysis. Nevertheless, it is worth noting that most (4/5, 80%) SHV/TEM producing isolates were ST636, and the isolates with CTX-M-14 and CTX-M-55 in combination were both ST69.

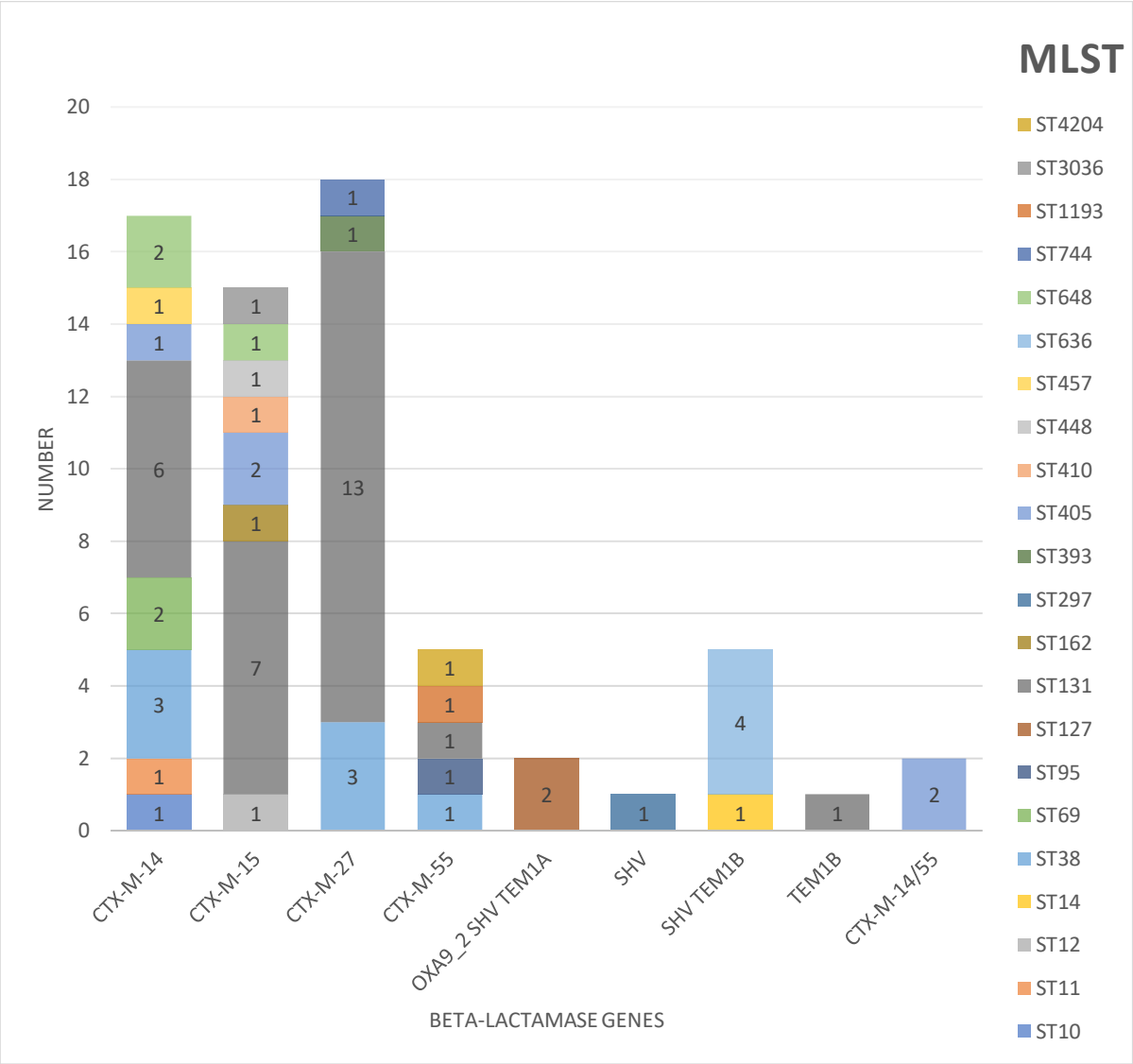



Figure 12: **MLST and beta-lactamase gene relationship**

### 3.2.4.3 *Other resistance genes*

The resistance genes for other antibiotic classes are illustrated in Figure 13. Genes encoding resistance to aminoglycosides, trimethoprim, sulphonamides, tetracyclines, macrolides, chloramphenicol, and quinolones were found. Notably, no complete colistin resistance genes were seen, though a partial *MCR-1* gene was found in one isolate. Only genes found with ≥95% coverage were considered (Supplementary Data, Appendix B).

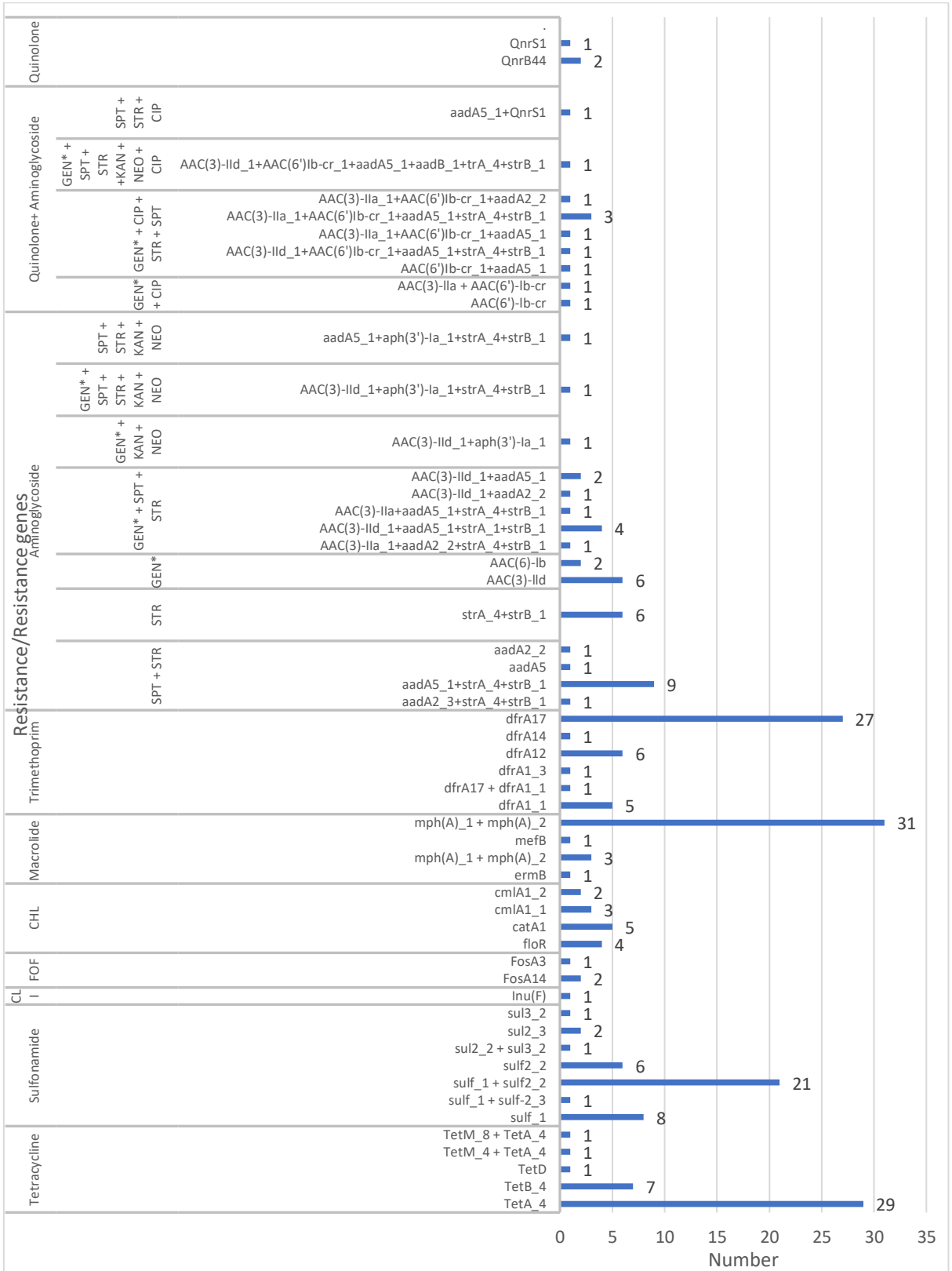
Of the 66 isolates, resistance genes against tetracyclines were found in 39 (59.0%), sulfonamides in 40 (60.6%), macrolides in 36 (54.5%), trimethoprim in 41 (62.1%), and aminoglycoside only in 38 (74.2%). Less commonly identified resistance genes included those to both quinolones and aminoglycosides in 11 (16.7%), quinolones only in 3 (6.1%), chloramphenicol in 14 (21.2%), fosfomycin in 3 (4.5%), and lincosamides in 1 (1.5%). Considering aminoglycoside modifying enzyme encoding genes individually, 29 (43.9%) isolates carried genes encoding resistance to gentamicin, 38 (57.6%) to spectinomycin and/or streptomycin, and 4 (6%) to kanamycin and neomycin. *strA* and *strB* streptomycin and spectomycin resistance genes always occurred together. Considering quinolone resistance, 14 (21.2%) isolates carried resistance genes to ciprofloxacin. Most (10/14, 71.4%) were *aac(6′)-Ib-cr* genes, with the remainder being *qnr* genes split evenly between *qnrS* and *qnrA*. Considering trimethoprim resistance, the *dfrA17* gene was the most common resistance determinant (27/41, 65.9%), followed by *dfrA1* (7/41, 17.1%) and *dfrA12* (6/41, 14.6%).

Genes encoding resistance to 3 or more classes of antibiotics, were found in 42/66 (63.6%) of isolates (151). This phenomena was found far more commonly in CTX-M type ESBL (39/57, 68.4%) than TEM-1B/OXA-9/SHV-12 carrying ESBLs (3/9, 33.3%) ($p<0.05\%$). The incidence of this phenomena did not differ greatly between different CTX-M types (CTX-M-14, 11/17 (64.7%); CTX-M-15, 10/15 (66.7%); CTX-M-27, 14/18 (77.8%); CTX-M-55, 2/5 (40%); CTX-M-14/55, 2/2 (100%). Of note, 8/10 (80%) *aac(6′)-Ib-cr* encoding isolates also encoded CTX-M-15.

#### 3.2.4.4 ***Comparison with Phenotypic Resistance***

Genotypic resistance was compared with phenotypic resistance, as determined by disc diffusion according to EUCAST guidelines in a previous study (78). Of the 14 isolates with genotypic quinolone resistance, all except one (encoding the *qnrS* gene) showed phenotypic quinolone resistance to ciprofloxacin. Phenotypic resistance to ciprofloxacin was seen in 51/66 (77.3%) isolates, 38 (74.5%) of which had no detectable quinolone resistance gene; mutations in the quinolone resistance determining region (QRDR) were not assessed. Of 29 isolates with genotypic gentamicin resistance, 6 (21%) were

phenotypically susceptible to gentamicin. On the other hand, 1 isolate showed phenotypic resistance and 1 isolate intermediate susceptibility to gentamicin despite no corresponding resistance gene. Out of 45 isolates encoding a *sul* or *dfrA* resistance gene, 8 (17.8%) were susceptible to co-trimoxazole. On the other hand, 2/39 (5%) showed resistance to co-trimoxazole despite no *sul* or *dfrA* genes being present. Data on phenotypic resistance to tetracyclines and fosfomycin was not available (78).




Key: spectinomycin (SPT); streptomycin (STR); gentamicin (GEN); ciprofloxacin (CIP); kanamycin (KAN); neomycin (NEO); clindamycin (CLI); fosfomycin (FOF); chloramphenicol (CHL)
*= Gentamicin, tobramycin and plazomicin

Figure 13: **Distribution of non-β-lactam resistance genes among ESBL-producing *E. coli* and the corresponding predicted resistance**

## 3.3 ISOLATION OF ESBL GENE-ENCODING PLASMIDS

ESBL encoding plasmids were able to be isolated from 45 of 66 (68%) ESBL *E. coli* isolates. Of these, 33 (73.3%) were isolated by conjugation with the sodium azide resistant *E. coli* J53 strain. Figure 14 shows an image of gel electrophoresis of putative transconjugants following K12-specific PCR to confirm the strain was J53, with Figure 15 showing the corresponding PCR products of multiplex II to confirm the presence of the ESBL gene from the donor strain in putative transconjugants (117,120). The remainder (12/45, 26.7%) were isolated by transformation of plasmid preparations from the clinical isolates into electrocompetent DH10B (10 isolates) or ST18 (2 isolates) *E. coli* strains. The presence of the ESBL gene from the donor strain in putative transformants was confirmed by PCR (Figure 16) (117).

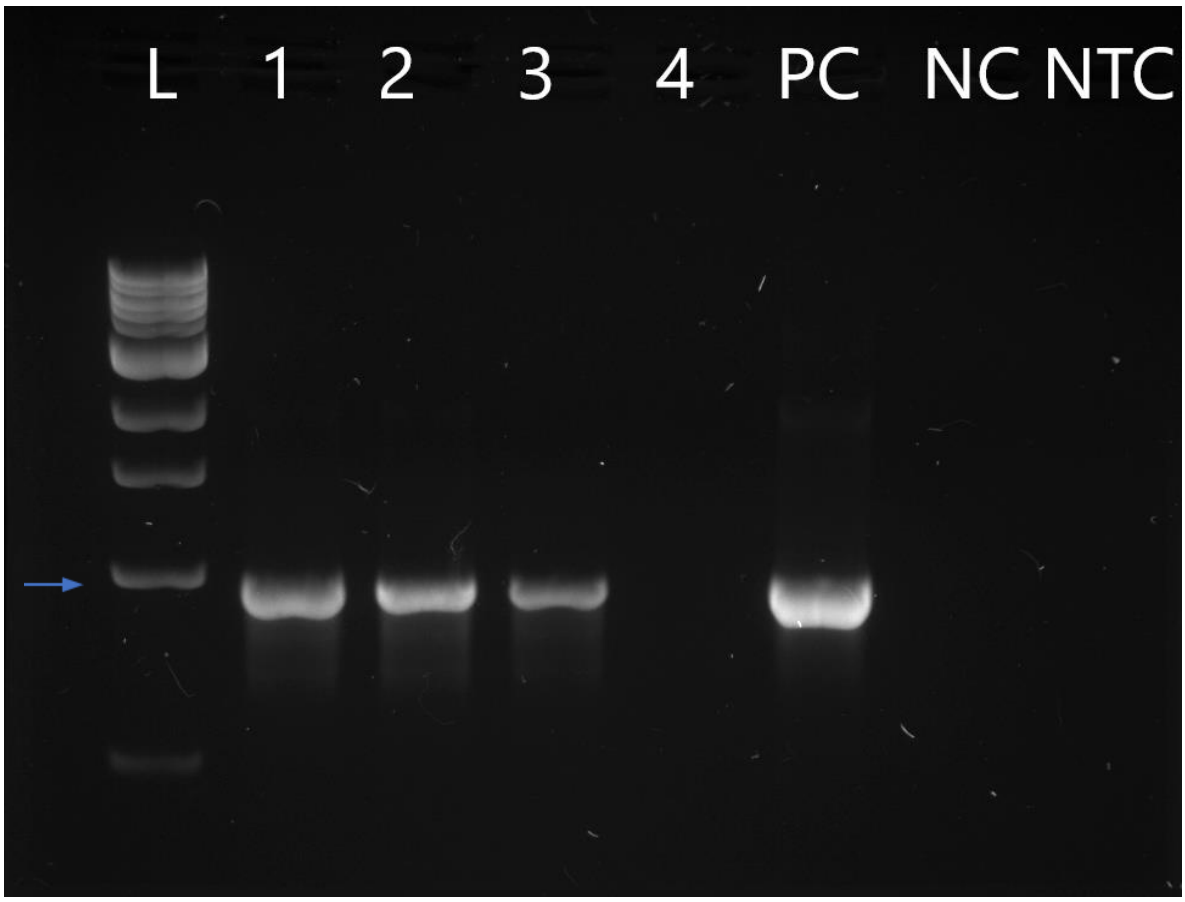



Figure 14: **K12 confirmation of putative transconjugants (conjugation).** Gel electrophoresis of PCR products of putative transconjugants to confirm the success of conjugation by the presence of K12 specific band **L:** 1Kb plus ladder : 100bp-12000bp (with 1000bp marked); **PC:** positive control: J53 strain with K12 specific band at 969 bp; **NC:** negative control ESBL *E. coli* clinical isolate- the donor isolate for the putative transconjugant in lane 1; **NTC:** no template negative control (DEPC water); **1-3:** successful transconjugants with K12 specific band at 969 bp- conjugation has occurred; **4:** unsuccessful putative transconjugant with no band, indicating failure of conjugation.

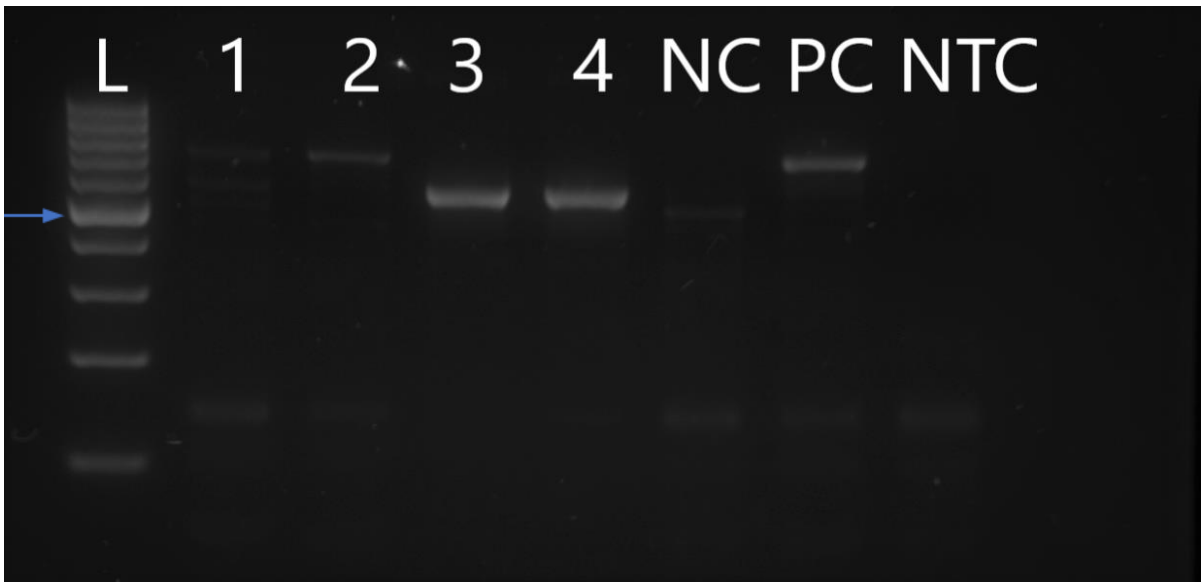



Figure 15: **Multiplex II ESBL gene confirmation in putative transconjugants (conjugation).** Gel electrophoresis of PCR products of multiplex II on putative transconjugants to confirm the presence of ESBL gene **L:** 100-1000 bp ladder (with 500 bp marked); **1** putative transconjugate with no band indicating that it is negative for CTX-M groups 1, 2, and 9 and that conjugation has failed; **2:** successful transconjugate with band at 688 bp, positive for CTX-M group 1, indicating successful conjugation; **3, 4:** successful transconjugates with band at 561 bp, positive for CTX-M group 9, indicating successful conjugation; **NC:** negative control J53 strain; **PC:** positive control CTX-M group 1 ST131 from ESR with band at 688 bp; **NTC:** No template negative control (DEPC water).

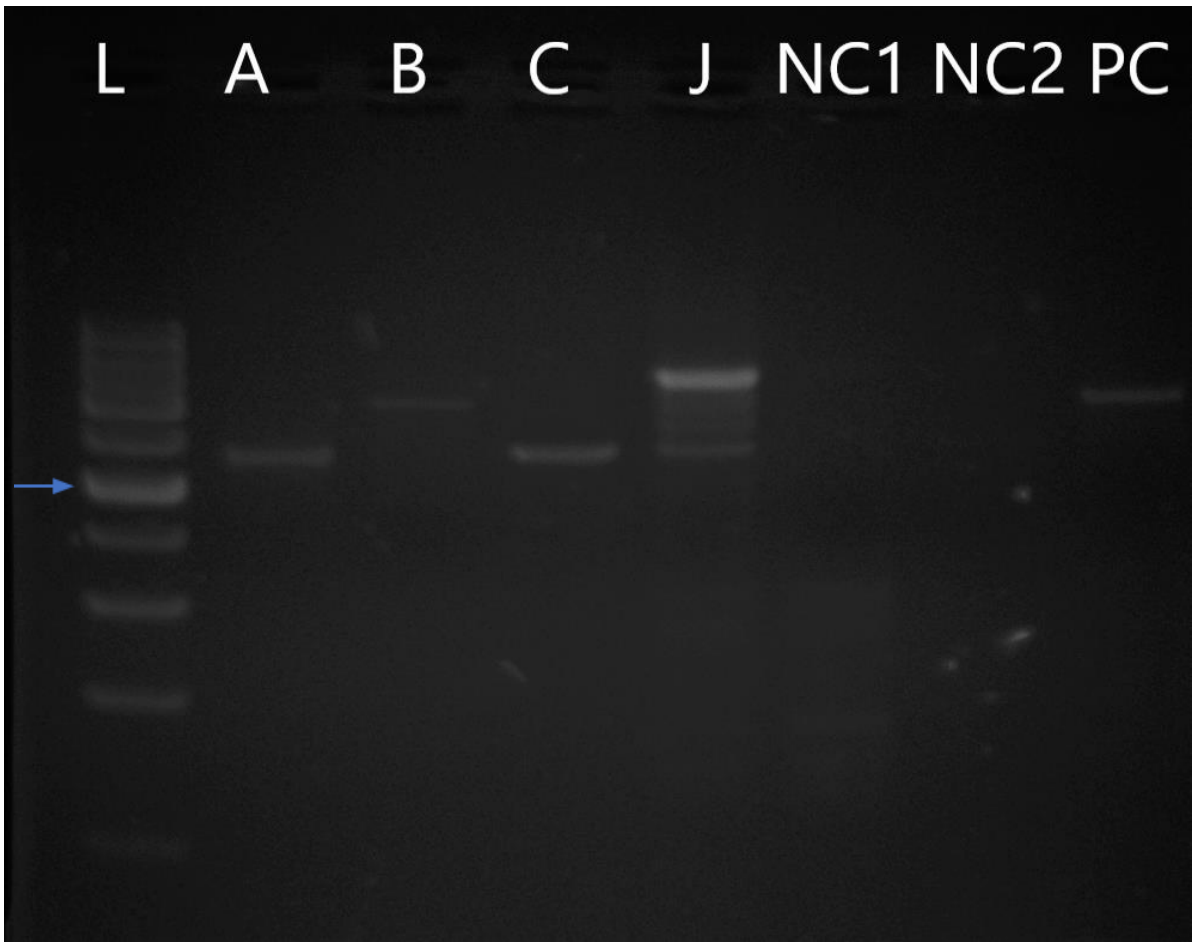



Figure 16: **ESBL gene confirmation in putative transformants (electroporation).** Gel electrophoresis of PCR products of multiplex II putative transformants to confirm the success of electroporation by the presence of ESBL gene specific PCR products **L:** 100-1000bp ladder (500bp marked); **A, C:** successful transformant with CTX-M group 9 specific band at 561bp - successful transformation; **B:** successful transformant with CTX-M group 1specific band at 688bp- successful transformation; **J**: putative transformant with multiplex I non specific laddering failed transformation **NC1:** negative control DH10B multiplex II; **NC2:** Negative control DH10B multiplex I; **PC:** positive control CTX-M group 1 ST131 from ESR 688bp

### 3.3.1 **Optimization of Conjugation Method**

Firstly, the optimal concentration of sodium azide and cefotaxime to use in the selective agar were determined by the agar dilution method. *E. coli* J53 failed to grow at a cefotaxime concentration of 1.5 mg/L but was able to grow at sodium azide concentrations of 150 mg/L to 175 mg/L within 24 hours; diminished growth was seen up to 275 mg/L after 48 hours incubation. All of the clinical isolates grew at a cefotaxime concentration of 1.5 mg/L. While most clinical strains failed to grow on sodium azide, one isolate grew at 200 mg/L. Therefore, a more resistant strain of *E. coli* J53 was

derived. This allowed a sodium azide concentration of 250 mg/L to be used for the conjugation experiments.

Secondly, the optimum ratio of donor to recipient strains for conjugation was determined. Ratios of 1:1, 1:2, 1:3 and 1:9 were assessed using a representative selection of six isolates. The only transconjugant elucidated (as confirmed by PCR) in this batch was from the 1:1 ratio. The optimum pore size of the Millipore membrane filter was determined by comparing 0.22 μm and 0.45 μm with a selection of 5 isolates on each. One isolate was able to be transferred with 0.45 μm that was negative with 0.22 μm. Five further test isolates run with 0.45 μm were all successful.

### 3.3.2 **Optimization of Electroporation Method**

Nanodrop spectrophotometer purity assessment of extracted plasmid DNA found A260/280 values ranged from 1.92- 2.15, and the A260/230 values ranged from 1.23-2.15 with a DNA yield ranging from 300-1300 ng/μL.

To determine the optimal plasmid concentration for electroporation, plasmids from 6 isolates were run at neat concentration (600-1300 ng/μL), 1 in 20 dilution, and 1 in 40 dilution. The optimum range was found to be between 40-70 ng/μL, with efficiency rate dropping either side (lower at 15-30 ng/μL and at neat concentration). The optimum electrocompetent cell volume was determined between 20 μL (~$4x10^9$ cells) and 40 μL (~$8x10^9$ cells). The optimum electrocompetent cell volume was 40 μL (~$8x10^9$ cells).

## 3.4 **MinION Sequencing**

### 3.4.1 **MinION Quality Control**

Nanodrop spectrophotometer purity assessment of extracted plasmid DNA found A260/280 values ranged from 2.08- 2.13 (before AMPureXP clean-up =2.14-2.18), and

the A260/230 values ranged from 2.48-2.52 (before AMPureXP clean-up =2.24-2.34) with a DNA yield ranging from 300-1300 ng/µL. Low percent agarose gel electrophoresis analysis of input DNA showed high molecular weight DNA and no degradation or shearing was detected.

The *EPI2me* quality analysis of the reads of all four MinION runs are summarized in Table 5. Run 1 and run 4 had the best quality scores. Run 4 had the longest read length, highest read number and highest yield average, though the quality score was lower than run 1. Run 4 used a new flow cell and run 1 used a flow cell that had only been used once previously. Freshly isolated plasmid was used in run 2 but this did not improve the quality score; though the read length average improved slightly, the read length mode was significantly lower. This compared to run 1 and run 4 where plasmids were stored between isolation, quality testing and analysis (2 days). Run 3 had the lowest quality score, read length, read number and yield. Run 3 plasmids were stored at -20$^{o}$C for 3 weeks prior to analysis.

Table 5: *EPI2me* quality analysis of MinION reads

| | Run 1 | Run 2 | Run 3 | Run 4 |
|---|---|---|---|---|
| Read length average (bp) | 12190 | 13045 | 10304 | 14070 |
| Read length mode (bp) | 4850 | 2791 | 1233 | 4416 |
| Quality score average | 11.53 | 9.705 | 9.5375 | 10.4775 |
| Quality score mode | 12.75 | 9.705 | 10.042 | 10.96 |
| Read number average | 1261 | 6089 | 647 | 2595 |
| Read number median | 1188 | 4326 | 561 | 2062 |
| Yield average (Mbp) | 20 | 50.94 | 6.86 | 37.11 |
| Run duration (h) | 5 | 12 | 12 | 6 |
| Plasmid storage time | 2 days | 0 days | 3 weeks | 2 days |

### 3.4.2 **Hybrid Assembly**

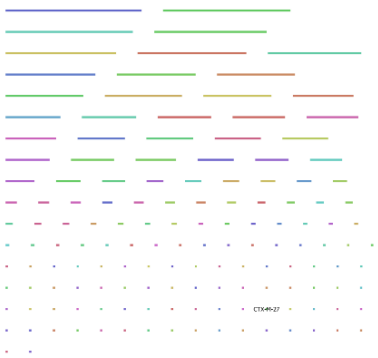
Different hybrid assembly methods, namely, *Miniasm/SPAdes*, *Miniasm/plasmid SPAdes,* and *Unicycler* were trialed on three isolates. The outputs for one example isolate is shown in Figure 17. The Illumina only assembly with *SPAdes* resulted in a fragmentated assembly with many divergent nodes where short read sequences failed to fully resolve the genome and any plasmids (Figure 17A and 17B); this illustrates the deficiency of short reads in resolving assembly continuity. A hybrid assembly using *Miniasm* to assemble long reads followed by *SPAdes* for hybrid assembly with short reads was assessed with a subset of three isolates; it showed a similar fragmented appearance to assembly of Illumina data alone (Figure 17C). The ESBL genes were located on small nodes with divergent paths; where small circularized plasmids were produced, they did not carry ESBL genes. While using the *'—plasmid'* option in *SPAdes* during hybrid assembly resulted in less fragmented assemblies, complete circularized plasmids carrying the ESBL gene in question were not obtained in any of the tested samples (Figure 17D). With the use of *Unicycler* complete plasmids carrying ESBL genes were obtained, though a highly fragmented assembly of other sequences (such as chromosome reads) was also present (Figure 16E). Prefiltering short reads with *bowtie2* removed these unwanted assemblies and resulted in assemblies of only the desired complete circularized plasmids (Figure 17F).

We were able obtain complete circularized plasmids from 25 isolates without using *bowtie2* filtering. The remaining 20 unresolved plasmids and an additional 5 already completed plasmids were run with the *bowtie2* prefilters. Using *bowtie2* prefilters resulted in a complete plasmid in 4 out of 5 (80%) plasmids that had been completed without the filter, with the remaining 1 giving a linear plasmid. Sixteen of the 20 (80%) previously unresolved plasmids were able to be completed using the *bowtie2* prefilter. Only one plasmid that was unable to be completed with *Miniasm* was able to be completed with *sed,* however plasmid type was unable to be identified.

Overall, we were able to obtain a complete circularized plasmid assembly in 41/45 isolates using the constructed pipeline. These four non-circularized plasmids were from

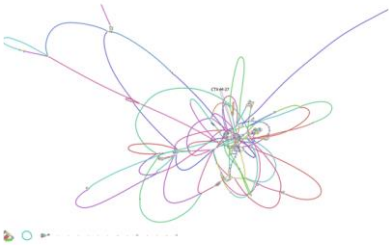
different MinION runs, therefore variation in sequence quality between runs was not the cause. However, sufficient information could still be obtained from 3 of 4 of these non-circularized plasmids to characterize the insertion sequences around the ESBL gene and plasmid type. Conversely, plasmid type was unable to be identified (with *PlasmidFinder*) in 3 other isolates (from MinION runs 1 and 2) despite completed, circularized plasmids having been obtained using the pipeline.

a)Illumina contigs

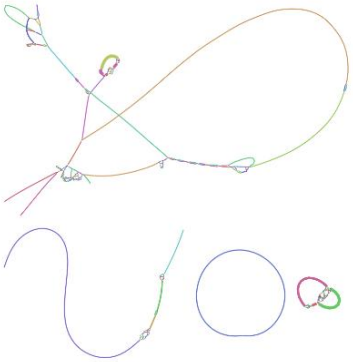


b)SPAdes- illumina only

c)Miniasm/SPAdes



d)Miniasm/plasmid SPAdes



e)*Unicycler* (ESBL gene carrying plasmid circled)

f)*bowtie2*/*Unicycler*


Figure 17: **Assembly of one test isolate with different tools**. The outputs of different tools used for hybrid assembly of one test isolate **a)** Illumina contigs from *Nullarbor*, short read length with CTX-M-27 on a 1767bp node **b)** Illumina only *SPAdes* assembly with CTX-M-27 on a 1767bp node in a fragmented assembly **c)** *Miniasm/SPAdes* hybrid assembly with CTX-M-27 on a 1667bp node in a fragmented assembly **d)** *plasmidSPAdes* with no blast hits for CTX-M-27 **e)** *Unicycler* output with CTX-M-27 on 105081bp plasmid (circled), in addition to branching assembly and another circular plasmid with no ESBL gene **f)** *bowtie2* followed by Unicycler with a complete circularized 105081bp plasmid carrying CTX-M-27, and no other fragmented sequences

### 3.4.3 **Beta-Lactamase Gene Type of Isolated Plasmids**

Of the 9 isolates which contained only one or more of TEM/SHV/OXA, a plasmid was able to be isolated from 6 (66.7%). Of 57 CTX-M encoding isolates, 39 (68.4%) ESBL-encoding plasmids were able to be isolated. While an ESBL-encoding plasmid could be isolated from most CTX-M-27 encoding isolates (16/18, 88.9%), it was only possible in 6/17 (42%) CTX-M-14 encoding isolates. Of the isolates encoding a group 1 CTX-M, a plasmid could be isolated from 9/15 (60%) CTX-M-15 encoding and 4/5 (80%) CTX-M-55 encoding isolates. A plasmid was obtained from both CTX-M-14/CTX-M-55 encoding isolates.

The ESBL genes encoded by the 45 plasmids that were able to be transferred are shown in Table 6. Whole genome sequencing identified all β-lactamase genes present in the isolates. With the MinION data, only the plasmid which was able to be conjugated or transformed in vitro was sequenced. Not all β-lactamases from the original isolates were present on the isolated plasmids. These other β-lactamases which were not transferred are also shown in Table 6. A CTX-M ESBL gene plasmid was transferred from 15 of the 16 (93.75%) isolates wherein CTX-M types and TEM/SHV/OXA type co-occurred. In the one exception a TEM-1B gene was isolated rather than CTX-M-14. There was also a plasmid isolated which carried both CTX-M-15 and TEM-1B, though it should be noted the ESBL determinant was CTX-M-15. In the two isolates where CTX-M-14 and CTX-M-55 co-occurred, a plasmid encoding only CTX-M-14 was isolated.

There were two cases of incongruence. One plasmid isolated was found to carry a SHV-183 ESBL gene, despite this gene not being identified in the Illumina WGS results (which identified CTX-M-14 and TEM-1B). In another instance, a plasmid which carried CTX-M-15 and TEM-127 was identified from an isolate which only carried OXA9/SHV-12/TEM-1A β-lactamase genes according to the Illumina WGS results. In both cases the anomalous gene (SHV-183 and CTX-M-15 respectively) was found in the raw MinION reads using the *EPI2me* resistance workflow. The WGS unfiltered Illumina reads were also assessed using CARD and the CGE ResFinder to ensure these anomalous genes were not missed due to differences in resistance gene databases; the genes were absent. The

genes were also absent in the *bowtie2* filtered Illumina reads. Furthermore, no β-lactamase gene was present in both sets of *bowtie2* filtered Illumina reads at all according to the CGE ResFinder. Thus, it appears these anomalies were caused by an error in the MinION analysis, probably due to contamination. For this reason, these two plasmids were excluded from later analyses.

Table 6: Beta-lactamase genes encoded by isolated plasmids

| Beta-lactamase gene encoded by isolated plasmids (MinION) | Beta lactamase genes present in original isolate (Illumina WGS) | Number |
|---|---|---|
| CTX-M-14[Y] | CTX-M-14 | 3 |
| | CTX-M-14/CTX-M-55 | 2 |
| | CTX-M-14/TEM1B | 3 |
| CTX-M-15 [Y] | CTX-M-15 | 4 |
| | CTX-M-15/OXA-1 | 3 |
| | CTX-M-15/OXA-1/TEM1B | 1 |
| CTX-M-15 [Y]/TEM-1B* | CTX-M-15/CMY/TEM1B | 1 |
| CTX-M-15 [Y]/TEM-127* [K] | OXA9/SHV-12/TEM1A | 1 |
| CTX-M-27 [Y] | CTX-M-27 | 12 |
| | CTX-M-27/OXA-1 | 2 |
| | CTX-M-27/TEM-1B | 2 |
| CTX-M-55 [Y] | CTX-M-55 | 1 |
| | CTX-M-55/TEM1B | 3 |
| SHV-12[Y] | SHV-12 | 1 |
| | SHV-12/TEM1B | 3 |
| SHV-183[X K] | CTX-M-14/TEM1B | 1 |
| TEM-81[+] | SHV-12/TEM1B | 1 |
| TEM-1B* | CTX-M-14/TEM1B | 1 |

*=broad-spectrum beta-lactamase phenotype
[+]=inhibitor resistant beta-lactamase phenotype
[X]=phenotype undefined
[Y] =ESBL phenotype
[K]= excluded from further analyses

### 3.4.4 **MLST**

The proportions of different sequence types from which a plasmid was able to be isolated was largely consistent with the proportions observed in the full set of isolates. ST131 made up the largest portion (17/43, 39.5%), followed by ST38 (4/43, 9.3%), ST405 (4/43, 9.3%) and ST636 (4/43, 9.3%). With the exception of ST14 and ST95, at

least one isolate of all sequence types present in the full set of isolates was able to be isolated by transformation or conjugation. Therefore, MLST type did not affect the ability to isolate an ESBL-encoding plasmid.

When considering the combination of MLST and ESBL gene type there were some differences. In the group of isolates from which CTX-M-14-encoding plasmids were isolated, only one (1/8, 12.5%) was from a ST131 isolate, despite 31.6% (6/19) of all CTX-M-14 encoding isolates being ST131. In contrast, a CTX-M-14-econding plasmid was isolated from all 3 of the CTX-M-14 positive ST405 isolates. A CTX-M-14 encoding plasmid was also isolated from 1 each of the other MLST types, except for the ST10 and ST38. In contrast, the proportions of different MLST from which a CTX-M-27 encoding plasmid could be isolated (11/16, 68.75%) was consistent with the distribution of MLST in all 66 isolates (the full set), with ST131 making up the majority (13/18, 72.2%). While ST131 was the most common MLST from which a CTX-M-15 encoding plasmid was isolated (3/9, 33.7%), it represented a smaller proportion than in the full set (7/15 46.7%). The three CTX-M-15-positives isolates from which a plasmid could not be isolated were ST648, ST405, and ST131. A plasmid could be isolated from all but two of the CTX-M-27 encoding ST131 isolates (11/13, 84.6%). Of the CTX-M-55 (excluding the co-occurring CTX-M-14 isolates), only the ST95 was not transformed. Figure 18 illustrates the diversity of MLST types from which an ESBL-encoding plasmid was isolated, with variation in the ESBL gene indicated. The two excluded isolates encoding SHV-183 and CTX-M-15/TEM-127 were from ST10 and ST127 MLST isolates respectively (not shown in figure).


*=broad-spectrum beta-lactamase phenotype
ᵗ=inhibitor resistant beta-lactamase phenotype
ᵞ =ESBL phenotype

Figure 18: **Plasmid encoded beta-lactamase genes and MLST of host *E. coli*.** ESBL genes encoded on plasmids were determined by hybrid assembly of MinION and Illumina plasmid reads. MLST of host *E. coli* were determined from Illumina WGS data.

### 3.4.5 **Plasmid Incompatibility Types**

Next, the plasmid incompatibility types of the isolated β-lactamase encoding plasmids were determined. Ten different incompatibility types were identified (Figure 19). Multi-

replicon plasmids were common in this set of isolates, with three different types of multireplicons identified: IncFIA/IncFII, IncN/IncX4, and IncFIB/IncFII. Multireplicon IncFIA/IncFII was one of the two most common incompatibility types in this set of plasmids, alongside IncFII (both n=10, 22%). IncI1 also made up a significant portion of the plasmid population (n=8, 18%). The plasmid incompatibility type of 4 of the isolates was unable to be identified.

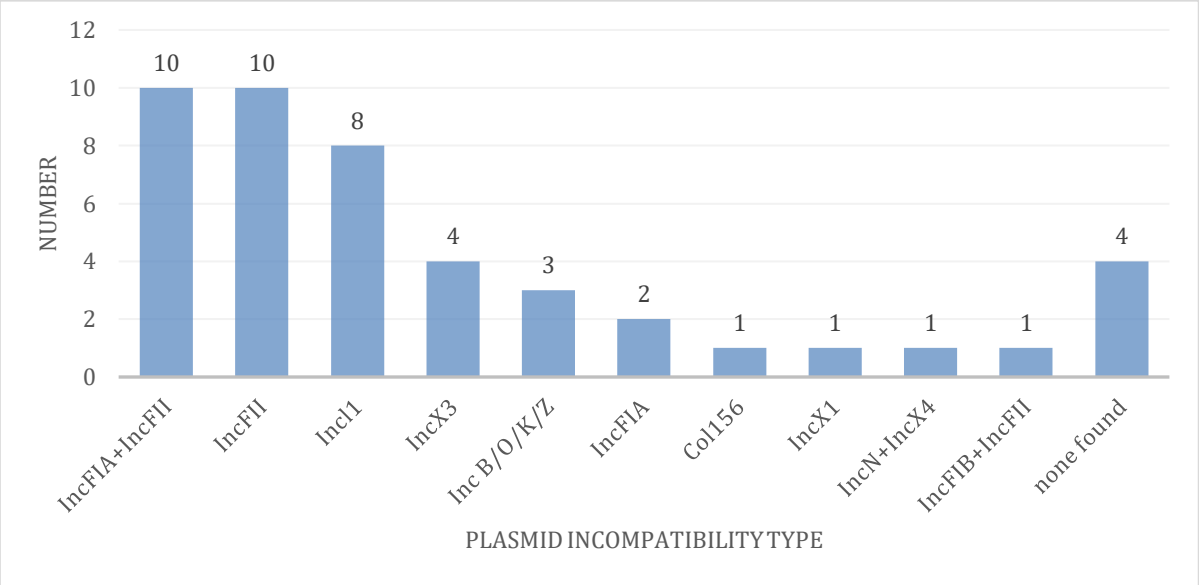



Figure 19: **Plasmid Incompatibility Types.**

Next, the plasmid incompatibility types were considered in context of the β-lactamase gene they carried and the MLST of the isolate from which they were isolated (Figure 20). There was a significant association between CTX-M-27 and IncF plasmids ($p < 0.05$). The CTX-M-14/ST405 cluster of 3 isolates was revealed to be carried on three different plasmid incompatibility types (IncFII, IncFIA, and IncB/O/K/Z). In addition, the large cluster of CTX-M-27/ST131 was revealed to be carried on multiple different plasmid types (IncFIA/IncFII, IncFII and IncB/O/K/Z), though 7 ST131 isolates with a CTX-M-27-encoding IncFIA/IncFII multireplicon plasmid were found. Only one other ST, ST405, carried an IncFIA/IncFII multireplicon plasmid, but that plasmid encoded CTX-M-15. In contrast, the SHV/ST636 isolates (n=3) all carried an IncX3 plasmid, and the CTX-M-27/ST38 isolates (n=3) all carried an IncFII plasmid. The plasmid incompatibility types of two of the three CTX-M-15/ST131 isolates were unable to be identified. Looking at

commonalities of carriage, all four IncX3 plasmids encoded SHV. The majority of both IncFII and IncFIA/IncFII plasmids encoded CTX-M-27 (both 7 of 10, 70%). IncI1 plasmids on the other hand carried mostly CTX-M-15 (50%, 4/8, 50%). The two excluded isolates encoding SHV-183/ST10 and CTX-M-15/TEM-127/ST127 were encoded on IncX3 and IncI1 plasmids respectively (not shown in figure).


*=broad-spectrum beta-lactamase phenotype
+=inhibitor resistant beta-lactamase phenotype
γ =ESBL phenotype

Figure 20: **Beta lactamase genes, MLST and plasmid incompatibility types of plasmid hybrid assemblies.**

### 3.4.6 **Insertion Sequences**

Insertion sequences surrounding ESBL genes on isolated plasmids were also able to be described. Insertion sequences found upstream of the ESBL genes in this study were IS*Ecp1*, IS*26*, IS*3000*, IS*Ehe3*, IS*1b*, and IS*1X2*, as well as transposases Tn*2* and Tn*A*. Downstream of the ESBL genes, orf*477*, IS*903D*, IS*26*, *iroN*, and IS*15DI* were found, as well as transposases Tn*2,* Tn*3,* Tn*A, tnpR*, and *deoR_TR*.

There were unambiguous differences in the insertion sequences around CTX-M genes and TEM/SHV genes. Insertions sequences such as IS*Ecp1*, IS*1X2*, *orf477*, IS*903D* and *iroN* appeared around CTX-M ESBL genes only, while IS*1b*, IS*3000*, IS*Ehe3*, *tnpR* and *deoR_TR* only appeared around TEM/SHV genes. The only commonalities were IS*26* and Tn*A*. There were also group specific differences. While IS*Ecp1* and IS*26* appeared upstream of the ESBL gene regardless of CTX-M group, Tn*3*-like IS*Ecp1* only appeared upstream of CTX-M-15. In contrast the downstream insertion sequences showed stark differences between CTX-M groups, with IS*903D* variants appearing only in group 9 CTX-M isolates, and *orf477* variants appearing only in group 1 CTX-M isolates.

The assemblage of these insertion sequences around ESBL genes is important, with more than one often appearing adjacent to the ESBL gene. As there are many insertion sequences (and combinations thereof) to consider, classification systems can ease the depiction of the diversity of the genetic context of the ESBL genes.

#### 3.4.6.1 ***Letter Allocation Classification System***

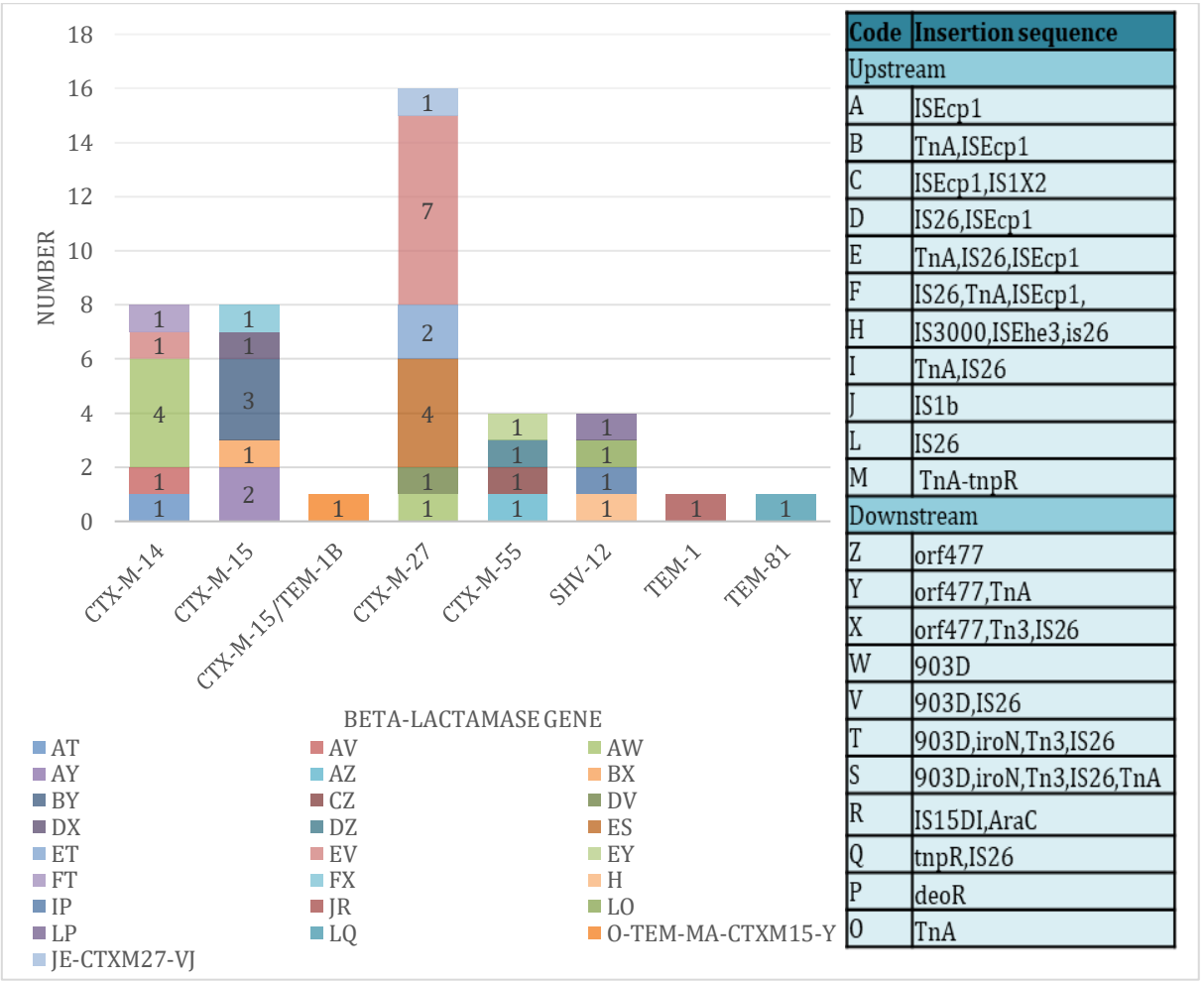
In order to easily compare and assess insertion sequence associations a classification system was developed that ascribes a character for each insertion sequence and transposon (Figure 21). Based on this classification system, Figure 21 shows the proportions of each insertion sequence found around each ESBL gene type in this study. Each CTX-M had a major (most frequently found) insertion sequences type, as well as some minor populations. The CTX-M-15 and TEM1B genes that were found next to each

other on the same plasmid were considered separately for the examination of insertion sequences.

We can observe that the insertion sequence combinations A-W (5'and 3' of ESBL gene) and E-V were found surrounding both CTX-M-14 and CTX-M-27 (both group 9 ESBLs, differing by one substitution- D240G), and that these were the only insertion sequence combinations to occur with more than 1 ESBL gene type. However, E-V occurred mostly in CTX-M-27, forming a large group of isolates (7/16, 43.75%), with only one instance in CTX-M-14 isolates; while A-W occurred mostly in CTX-M-14 (4/8, 50%), with only one instance in CTX-M-27 isolates. Excluding A-W, all four other CTX-M-14 isolates had different surrounding insertion sequences. In CTX-M-27 there was a second smaller group with the same surrounding insertion sequence E-S (4/16, 25%).

All the CTX-M-55 were different from each other in terms of insertion sequences. CTX-M-15 isolates had high diversity with only B-Y (3/8,37.5%) and A-Y (2/8, 25%) occurring in more than one isolate; these differed only in the presence of Tn*2* upstream of IS*Ecp1*.

All SHV-12 isolates had different surrounding insertion sequences (H-, L-P, I-P, L-O). In one of these (H-), the only insertion sequence identified occurred upstream of the SHV-12 gene. The TEM genes (TEM-1B, TEM-81) also had differing surrounding insertion sequences (J-R, O-M, L-Q).



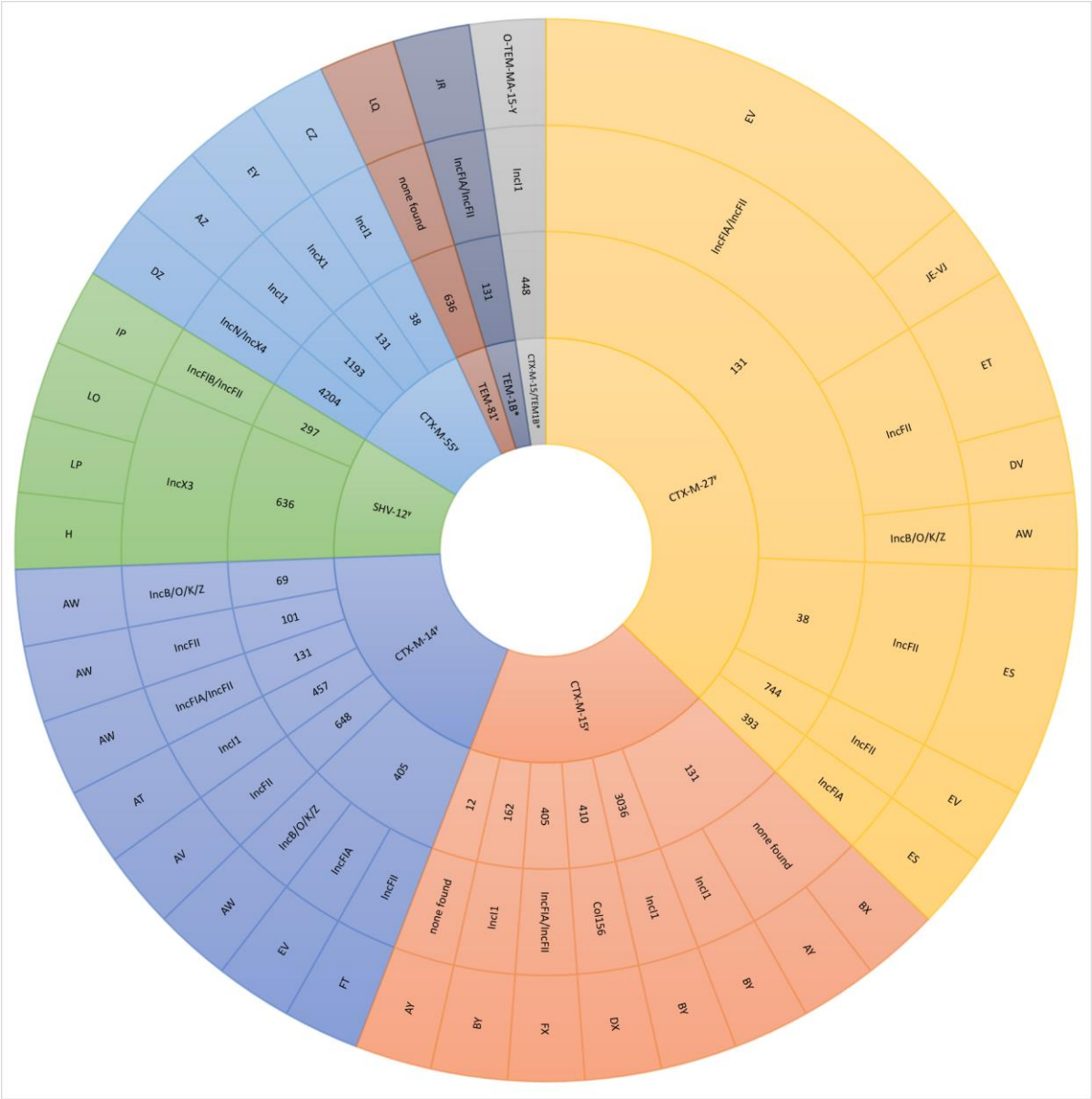

| Code | Insertion sequence |
|---|---|
| Upstream | |
| A | ISEcp1 |
| B | TnA,ISEcp1 |
| C | ISEcp1,IS1X2 |
| D | IS26,ISEcp1 |
| E | TnA,IS26,ISEcp1 |
| F | IS26,TnA,ISEcp1, |
| H | IS3000,ISEhe3,is26 |
| I | TnA,IS26 |
| J | IS1b |
| L | IS26 |
| M | TnA-tnpR |
| Downstream | |
| Z | orf477 |
| Y | orf477,TnA |
| X | orf477,Tn3,IS26 |
| W | 903D |
| V | 903D,IS26 |
| T | 903D,iroN,Tn3,IS26 |
| S | 903D,iroN,Tn3,IS26,TnA |
| R | IS15DI,AraC |
| Q | tnpR,IS26 |
| P | deoR |
| O | TnA |

Figure 21: **Surrounding insertion sequences by letter allocation system**

## 3.5 Overall Genetic Diversity

The overall genetic diversity can be considered by taking into account the ESBL gene, MLST, plasmid replicon type, and the insertion sequences surrounding the ESBL gene (Figure 22). This revealed that the three SHV/ST636/IncX3 isolates had different insertion sequences around the ESBL genes (L-O, L-P, and H-). Similarly, the three CTX-M-15/ST131 isolates had three different surrounding insertion sequences (B-X, A-Y, and B-Y). The CTX-M-27/ST131/IncFII isolates had one of two different surrounding insertion sequences (E-T and D-V). In contrast, the CTX-M-27/ST131/IncFIA+IncFII isolates were all EV class insertion sequences, though one of the isolates was also bracketed by IS1b (JE-VJ). The ST38/IncFII all had ES insertion sequences.

It is also apparent that there were some commonly occurring insertion sequences surround certain CTX-M-genes, that were in fact on plasmids of different incompatibility types and in isolates with different MLST profiles. For example: CTX-M-14 with A-W was on 4 different plasmids, carried by isolates with 4 different sequence types. CTX-M-15 with B-Y were all on plasmids of the same incompatibility type (IncI1), but were carried by isolates of different sequence types. The incompatibility type of the plasmids encoding CTX-M-15 with A-Y surrounding insertion sequences was unable to be identified in two cases and IncI1 in the other, but they were all from isolates of different sequence types.




*=broad-spectrum beta-lactamase phenotype
ᵗ=inhibitor resistant beta-lactamase phenotype
ᵞ =ESBL phenotype

Figure 22: **Full Genetic Diversity of ESBL *E. coli***

Finally, the core genome (Figure 23) and pan genome phylogenetic trees (Figure 24), obtained from WGS analysis, were annotated with all available information on ESBL gene, sequence type, plasmid incompatibility type, and surrounding insertion sequences (149). The isolates which were identical across all four assessed criteria are highlighted. While there is a cluster of six (6/66, 9.1%) identical isolates (ST131 isolates with plasmid CTX-M-27/E-V/IncFIA/IncFII), and two smaller clusters (two [2/66, 3.0%] ST131 isolates with plasmid CTX-M-27/E-T/IncFII; three [3/66, 4.5%] ST38 isolates with plasmid CTX-M-27/ES/IncFII), the majority of isolates appear to be genetically diverse.

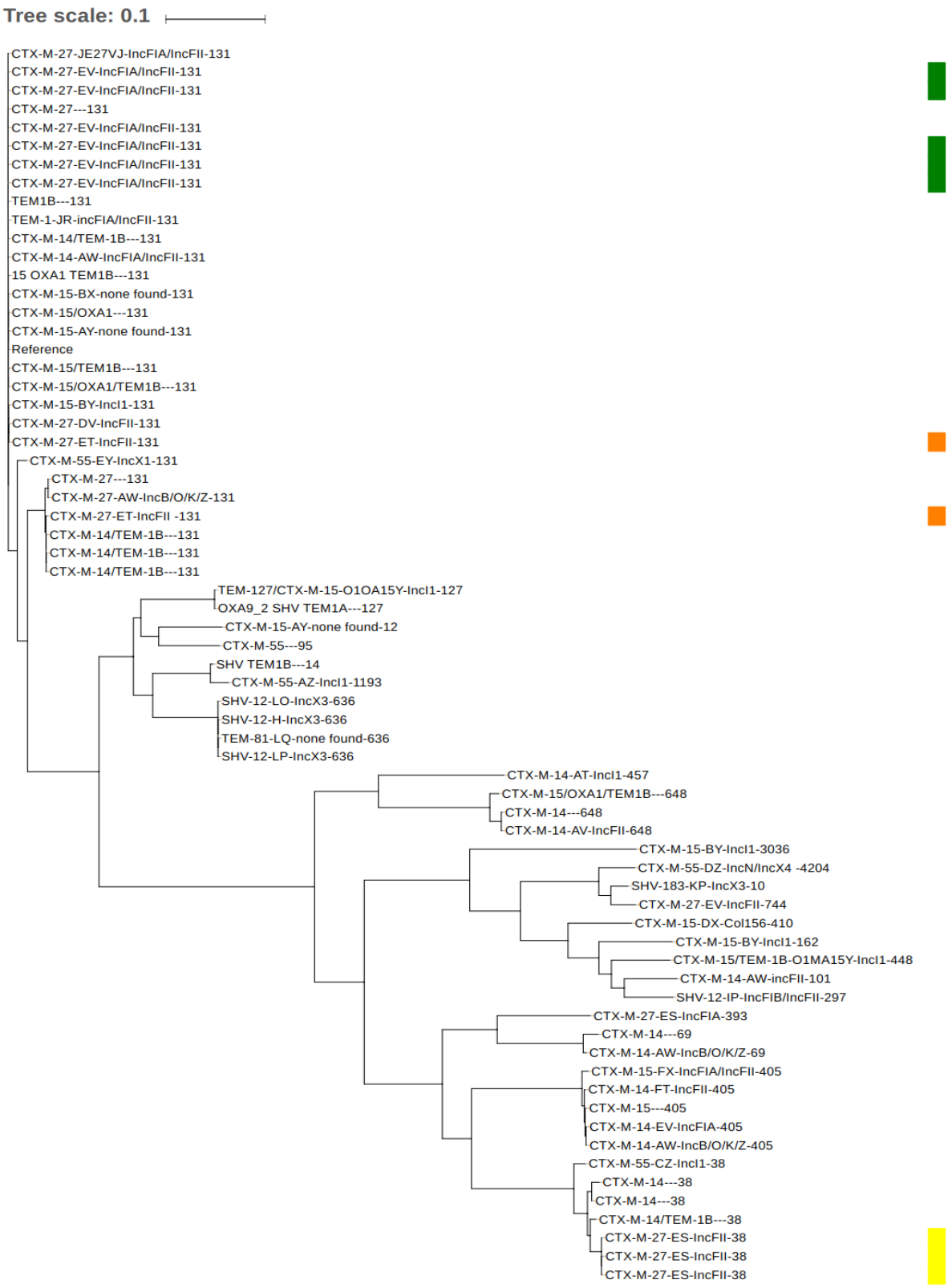



Figure 23: **Core genome SNP phylogenetic tree labelled with ESBL genes and genetic context.** Identical isolates in terms of gene, insertion sequences, plasmid incompatibility type and MLST are marked with the same colour. Graph generated and annotated using iTOL online (version 3.0) reference (149)

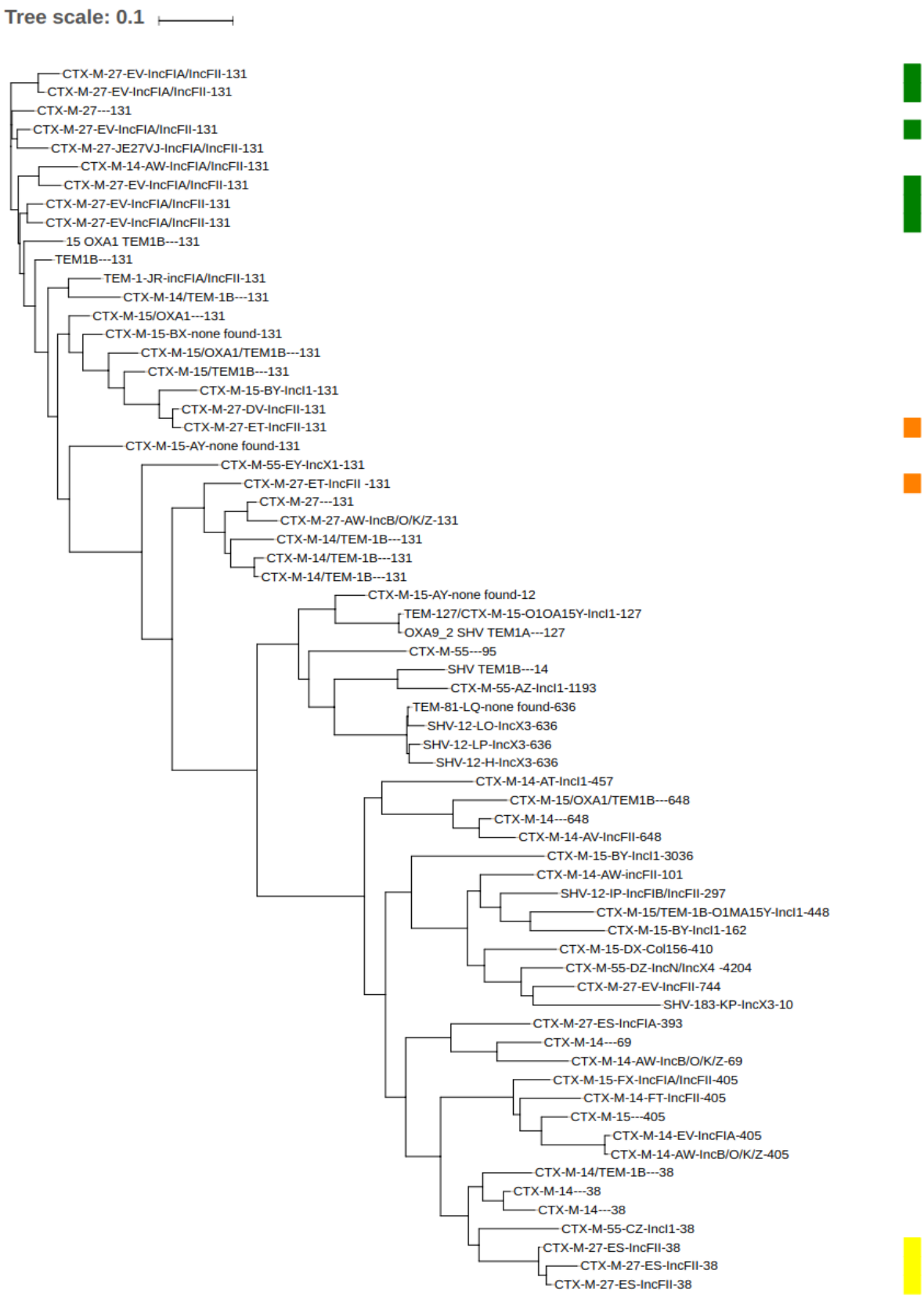



Figure 24: **Pan genome phylogenetic tree labelled with ESBL genes and genetic context.** Identical isolates in terms of gene, insertion sequences, plasmid incompatibility type and MLST are marked with the same colour. Graph generated and annotated using iTOL online (version 3.0) reference (149)

# 4 DISCUSSION

In this study, I characterized the genetic context of resistance and the phylogenetic relationship of ESBL *E. coli* isolated from urine from Otago between February 2015 and January 2016. A diverse array of ESBL genes, sequence types, plasmid types, and insertion sequences were found; suggesting that the recent increase in incidence in uropathogenic ESBL *Escherichia coli* in Otago is due to multiple introductions.

## 4.1 CLINICAL ISOLATES

All non-duplicate ESBL *E. coli* isolated by SCL from the Otago community and from Dunedin Hospital that were available were characterized in this study. That is, the whole population (n=82) of that year rather than a random selection. However, some isolates (n=13) were missing from storage. This represented only 15.9% of isolates and no systematic bias was apparent. In a previous study, three cryopreserved isolates were found not to be ESBL producers as had been reported; two were presumptive AmpC β-lactamase producers and one was susceptible to cefpodoxime (78). This may have been due to sampling error during storage in the presence of a double population of ESBL/non-ESBL bacteria. Therefore, the isolates in this study are an accurate representation of the diversity of ESBL *E. coli* causing urinary tract infections in the Otago community in 2015.

## 4.2 SPECIES IDENTIFICATION

Whole genome sequencing confirmed that all isolates were *E. coli*. *Kraken* (0.10.5) was the bioinformatic tool used for species identification in the *Nullarbor* pipeline (131,152). Kraken queries a look up table in a constructed database for exact k-mer matches and stores the lowest-common ancestor (LCA) if a k-mer is found in two or more taxa (152,153). The observed identity classification percentages (range 33.24%-82.23%,

median 72.69%) for the first match are expected for *Enterobacteriaceae* using *kraken* (152,153). The genera in the family *Enterobacteriaceae* have a high sequence similarity, posing difficulties for any classifier (152,153). Though the identity classification percentage was low in many of the isolates, the second match had very low identity percentage (<2.28%) comparatively in all isolates, making the identification of *E. coli* acceptable. The isolates had also all been confirmed as *E. coli* in a previous study using MALDI-TOF MS (78).

## 4.3 MLST AND PHYLOGENETIC ANALYSIS OF ISOLATES

The core genome is defined as the genes common to all strains of a species (39). The accessory genome represents the genes not common to all strains of a species (39). The pan genome refers to the entire set of genes - the combination of both the core genome and accessory genome (39). The *Nullarbor* bioinformatic pipeline uses *Snippy* and *FastTree* to construct a core genome phylogenetic tree. *Snippy* finds SNPs between the assigned reference (in our case an ST131 CTX-M-15 ESBL *E. coli*) and all the isolates, producing an alignment of the SNPs on genomic positions that are present in all samples (known as "core sites") (134). *FastTree* is used to construct the phylogenetic tree (131,135). The *Nullarbor* bioinformatic pipeline uses *Roary* to construct a pan genome phylogenetic tree with the input being annotated contigs (136). *Roary* takes into account full sequences of coding regions, and clusters isolates based on gene presence in the accessory genome (includes putative coding genes absent from one or more isolates), weighted by cluster size. There were a few (6-7) clusters evident in the core genome phylogeny tree, while the pan genome phylogeny tree showed several (9-10) smaller clusters. The variety in the pan genome phylogeny suggests some variation in the accessory genome. There are subgroups evident in the pangenome tree of the closely clustered isolates seen in the core genome phylogeny, suggestive of a shared accessory genome, and thus in the encoded mobile genetic elements. In a previous study, subgroups in the accessory genome of isolates that are closely related in core genome (with the same MLST [ST131]) has been described, with certain IncF plasmid allele

types associated with different ST131 sublineages, and further subgroups of IncF plasmid allele types (F29:B10 and B1) in the sublineage of H22 (154).

*Nullarbor* uses *mlst* which scans contig files against traditional (autodetected) PubMLST typing schemes using the NCBI Basic Local Alignment Search Tool (BLAST+) blastn database to identify the MLST of each isolate (131). The advantages of using MLST are that it is well established and can therefore be used for standardised comparisons across studies, can be used in both WGS, older methods, and in future genome sequencing methods, and allows for direct interpretation (74).

Worldwide, many *E. coli* sequence types have been identified to produce an ESBL (2,15). In this study the most common ST identified was ST131 (42.4%). Though it did not represent the majority of STs as seen in many countries internationally, the proportion found in this study is within reported ranges of 23-64% (3,4,15,16,25). This percentage was also lower than that seen nationwide in the 2013 ESR survey (54.5%, 122/224) and the 2016 ESR survey (60.1%, 176/293). The ESR surveys describe the nationwide hospital and community ESBL populations in New Zealand over a 1-month period. ST131 has been implicated in the rise in ESBL resistance worldwide, and is particularly important in ESBL resistant urinary tract infections due to their increased biofilm formation ability and common resistance to fluoroquinolones (3,4,14-17,19).

The second most common ST in this study, ST38 (7/66= 10.6%), was found at a similar frequency in the 2016 ESR survey (6.8%, 20/293); only ST131 was tested for in the 2013 ESR survey (13). ST38 has previously been identified as a MDR high risk clone, which is predominantly associated with urinary tract infections (14). This clone has been observed as one of the most common MLSTs in countries such as Saudi Arabia (Jeddah [2018] and Riyadh [2015]), Germany (2014), Korea (2012), Canada (2000-2010), Denmark (2017) and to a lesser extent from 2000-2010 in Norway, Japan, Spain, Turkey, France, and North America (14,19,22,36,73,155-159). Asymptomatic human faecal carriage has been reported, and it has also been observed in environmental bacterial populations isolated from animals (including fish), vegetables, and water, in

countries including Switzerland, Germany, the Netherlands, the United Kingdom, India, Thailand, and Vietnam (19,33,160-162).

ST405 (7.6% and third most common in this study, 2.4% and the seventh most common in the 2016 ESR survey) and ST648 (4.5% in this study, 2% in the 2016 ESR survey) were the only other MLST types that appeared in more than two isolates in both studies. ST405 has been important in the spread of CTX-M worldwide in the past (19). ST405 has also been linked to fluoroquinolone resistance and has been isolated in Norway, Kuwait, Spain, Switzerland, Belgium, France, Japan, Korea, Thailand and Australia; it has often been identified as one of the more common MLST types in ESBL *E. coli* (19,163-165).

Both my study and the 2016 ESR survey showed a high degree of diversity in STs. The ST of 12.2% of isolates (36/293, 25 different STs) in the ESR survey and 28.8% of isolates (19/66, 17 different STs) in this study were shared by less than or equal to 2 isolates. The higher rate in this study likely represents under-sampling due to the smaller sample size. There were 22 different non-ST131 STs in this study and 33 different non-ST131 STs identified in the ESR 2016 study. Some STs, which had low incidence in this study including ST69 (2/66, 3.0%), ST73 (not isolated), ST95 (1/66, 1.5%) and ST1193 (1/66, 1.5%), have been reported with high incidence in international regions in Europe (United Kingdom, Germany, Netherlands) and USA (California) (2). ST1193 was the third most common MLST in the 2016 nationwide ESR study (6.8%, 20/293). In summary, the dominance of different clones varies between countries and different regions with countries, with the exception of ST131.

## 4.4 PLASMID ISOLATION

Both conjugation and transformation were used for isolation of the ESBL-encoding plasmid. The conjugation method better demonstrates the transferability of these plasmids in the environment, as opposed to electroporation. The efficiency of the

conjugation method should also be considered. Cefotaxime was chosen to select ESBL transconjugants as CTX-M type ESBLs (most common in this set of isolates) display a higher affinity for cefotaxime than ceftazidime, and in the previous study with these isolates, 15 isolates did not display synergy with the ceftazidime/ceftazidime + clavulanic acid double disc diffusion test (5,10,13,15,18,19,78). The filter mating method was chosen as studies have shown this method to have higher transfer rates than liquid conjugation and solid phase conjugation (118,166). In agreement with these studies, previous attempts with liquid conjugation with this set of isolates gave low transfer rates (17/66, 26%), while the filter mating conjugation method was nearly twice as successful (33/66, 50%) (78). A previous study by Sasaki et al. found a higher frequency of conjugal transfer with a 0.45 μm filter compared to a 0.22 μm filter (167). The use of a 0.45 μm filter in my study allowed the transfer of 1 isolate (out of 5) that was not able to be transferred with a 0.22 μm filter. Using a larger sample size would have allowed a better comparison. A further five isolates run with the 0.45 μm filter were all successfully conjugated, so this filter was used for all isolate conjugation experiments. The study by Sasaki et al. also found that passing sterilized water or PBS through the filter under reduced pressure (about-50cmHg) before filter incubation increased the transfer rate (167). The use of this technique may have improved the transfer rate but was not attempted due to constraints of time, equipment, and cost. The large size of the plasmids (40-150 kb) may have affected transferability under laboratory conditions (127). A plasmid was able to be isolated from isolates of nearly all MLST types (except two). All MLST types carried transferable plasmids, and none in particular were associated with highly transmissible plasmids.

Successful electroporation was achieved in 12/33 (36.6%) plasmid isolates where a plasmid could not be transferred by conjugation. Electroporation was chosen over chemical treatment due to its higher efficiency (127). Methodology was refined during the project. DH10B was chosen as the it showed higher efficiency for large plasmids compared to other strains (H5a, XL-1 BlueMR, JS4 and HB101) in a previous study (127). The optimal electrocompetent cell volume was determined to be 40 μL ($8x10^9$ cells) in this study compared to 20 μL ($4x10^9$ cells), though only a small subset (6 plasmids) was used to test this. This was done as a previous study has shown cell concentration has an effect on transformation, with the peak being 5-$6x10^9$ cells (168). Similarly, the optimal

plasmid DNA concentration has previously been determined to be 10 to 40 pg (168). In my study 40-80 ng/µL was determined to be the optimal concentration, but again this was with a limited number of dilutions and test plasmids. The default settings of 1.8 kV (18 kV/cm), 200 ohm, 25 µF in a 0.1 cm cuvette were used for electroporation. I did not attempt to determine the optimum electroporation conditions due to time constraints. In pervious studies, the optimum conditions for plasmids >80 kb was determined to be 13 kV/cm (field strength), 100 ohm (resistor), 25 µF (capacitor) in a 0.2 cm cuvette (127,169). With 0.1 cm cuvette the optimum field strength was 20 kV/cm in one study (169).

The modified alkaline lysis method was used to isolate plasmids from clinical isolates for electroporation, and from transformants and transconjugants for sequencing on the MinION. Nanodrop spectrophotometer was used to assess the purity of the extracted plasmids. The A260/280 values for both experiments were higher (1.92- 2.18) than the benchmark for purity (1.8) even after cleaning with Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK). The A260/230 values of plasmids extracted for electroporation were lower (1.23-2.15) than the purity benchmark (2.0-2.2). The A260/230 values of plasmids extracted from the transformants and conjugants were higher (2.48-2.52) than the purity benchmark. Contamination with an extraction reagent such as phenol or an acidic pH normally lowers the A260/280 ratio (124,125). Higher A260/280 ratios are seen in basic solutions (124,125). The ratio is also dependent on the composition of the nucleic acid, as the different nucleotides display varying A260/280 results (124,125). A260/230 values were higher than the respective A260/280 values, which is expected in “pure” nucleic acid. The A260/230 ratios lower than 2.0 that were observed in the electroporation plasmids may be due to the presence of co-purified contaminants (124,125).

## 4.5 MinION and Bioinformatics

MinION based sequencing is a new technology, with improved techniques and bioinformatic software being developed rapidly (94). This has presented a challenge for ideal experimental design and resulted in each batch being processed differently. For example, the *MinKNOW* software was updated from version 2.0 to 2.1 during the study. *MinKNOW* 2.0 included a new graphical user interface as well as improvements such as a progressive unblock mechanism that improves removal of problematic strands by using progressive flicks instead of only a single flick, a dynamic voltage control that tracks current to adjust voltage as needed rather than in fixed steps at set time intervals to better stabilize run conditions throughout the experiment (170). *MinKNOW* 2.1 changed the mux scan (ranking of pores by quality) from a latching mechanism to a time based assessment, increased 'chunk' size (sectioning of raw signal data for base calling), and an improvement to the unblock mechanism to prevent incorrect ejections (171). These updates may have improved yield and quality, thus comparison between batches is difficult.

There were also differences in pre-analytical steps that may have resulted in inter-batch differences. From the variation in the four Minion runs that were performed, it appeared that the sequence quality decreased with storage of plasmids for longer periods of time, suggesting plasmid degradation. The run with plasmids that had been stored for the longest (run 3: stored 3 weeks at -20$^{o}$C) had the lowest average yield, read length, and quality score, though not the lowest quality score mode. Yet it appeared that storage at -20$^{o}$C for 2 days (run 1 and run 4) did not affect sequence quality compared to plasmids that were never frozen and run on the day of isolation (run 2), with both run 1 and run 4 having a higher quality score, and run 4 having a higher read length average than run 2. However, the age of the flow cell may have affected the results obtained with the fresh plasmids, with run 4 being run on a new flow cell, and run 1 being run on a flow cell used only once before.

In addition, only one batch was run with each condition, so more replicates would be

required to confirm these observations. Even so, rerunning on fresh flow cells or with plasmids with shorter storage time may have improved yield and quality. Inconsistencies between runs have been described in previous studies, with one study reporting differences between runs in the same laboratory as well as between two different laboratories, though the rapid kits were better than others in terms of yield consistency, especially with simpler library preparation, though it was difficult to predict yield (94).

Different run times were also used between batches. Two batches (run 1 and 4) were sequenced for 5-6 hours, while the other two batches were sequenced for 12 hours (run 2 and 3). The 12 hour runs were necessary as DNA library preparation was carried out during the day and sequencing performed overnight. Batches 2, 3 and 4 were not basecalled live to prevent data loss from lost connections; instead raw reads (.fast5) were later basecalled with *Albacore* to produce .fastq files. The first batch was base called live with *MinKNOW* which outputs only fastq files, so fast5 files for were not available to later analyse with the *albacore* pipeline used for the other isolates.

Yield (in Mbp) was more dependent on read quality (measured by quality score) than run duration. The first run with an average quality score of 11.53 had a yield of 20Mbp in 5 hours (4 Mbp/h), the second run with an average quality score of 9.705 had a yield of 50.94 Mbp in 12 hours (4.25 Mbp/h), the third run with an average quality score of 9.535 had a yield of 6.86 Mbp in 12 hours (0.57 Mbp/h), and the fourth run with an average quality score of 10.4775 had a yield of 37.11 Mbp in 6 hours (6.19 Mbp/h). Run 3 had a lower yield than run 1 and 4 despite being run for longer, and run 4 had a higher yield per hour than run 2. The fact that most pass reads are generated early in the run has been previously established (99). The maximum run time of 48 hours was not utilized. A previous study reported that only 31 pass reads were generated in the final 12 hours of a 48 hour run, with 50% of pass reads generated in the first 6 hours, and 90% of pass reads generated at 12 hours, suggesting that running for longer than 12 hours is unnecessary (99). However, in the same study, using only pass reads from the first 6 hours was less accurate than using 6 hour pass reads in addition to selections of pass reads from 6-12 hours. Therefore, it is possible that extending run 1 and 4 beyond

5-6 hours may have improved plasmid assembly. In contrast, the lower quality of reads beyond 6 hours may have affected the quality score average of run 3 and 2, both sequenced for 12 hours.

Despite differences in yield and quality score between runs, most plasmids were able to be completed using *bowtie2* and *Unicycler*, suggesting that the quality and yield of sequence data were sufficient for the scope of this project. All isolates where the plasmid type was not found were from runs 1 and 2, which had a quality average of 11.53 and 9.71 respectively. These quality scores were higher than the quality score of run 3 (9.5375) There were incomplete plasmids from each of the 4 runs. Therefore, variation in individual plasmid and library preparation quality may have been more important than variations between runs (keeping in mind there were only 4 runs) for insufficient sequencing data. In addition, large repeat structures may still cause problems when long reads are used if these repeats are longer than the MinION reads (83,100,102,109). Underrepresentation of homopolymers (stretches of identical bases) still pose a problem; though in one study *Miniasm* performed better than *Canu* in this regard, only the use of a polishing software such as *Nanopolish* sufficiently improved homopolymer detection (83,100,102,109).

### 4.5.1 **Basecalling**

The base caller used in this study, *Albacore*, has been shown to have greater speed and read accuracy than other base callers such as *Guppy* and *Chiron* in a comparison by Wick et al (172,173). Other advantages of *Albacore* include multi-threading support, the ability to base call fastq files, and a command line interface (109,174). However, *Scrappie* has similar speed and is more accurate in resolving homopolymers (174).

The version of *Albacore* used in this study includes demultiplexing. *Porechop* was also used to remove adaptors. *Porechop* also includes demultiplexing, and when *Porechop* disagrees with *Albacore* the read is discarded. This provides a two-step confirmation

which can decrease the likelihood for misfiled barcodes but can also result in barcodes being erroneously discarded. As *Porechop* was not trained on the rapid barcoding kit used in this study, its accuracy in demultiplexing may be lower than applications like *Deepbinner* which was trained and designed for this kit (174). However, *Deepbinner* requires raw reads (.fast5 files) which were not available for the first run.

### 4.5.2 **Miniasm/SPAdes**

I first used *Miniasm* to construct long read assemblies due its faster turnaround time compared to other long read assemblers such as *Canu. Miniasm* can take as little as 2 minutes, whereas *Canu* takes approximately 2 hours to complete 1 assembly (99,109). However, without a polishing step, with *nanopolish* for example, *Miniasm* is less accurate than *Canu.* An error correction step would have increased the accuracy of the *Miniasm* long read assembly, but this would also have significantly increased the turnaround time. For example, using *nanopolish* could increase the turnaround for one assembly to over 24 hours, even though *nanopolish* can improve consensus accuracy to 99.9% (98,99,139,172). I hypothesised that a polishing step or the use of the more accurate *Canu* was not required as *Illumina* reads would essentially be used for error correction in hybrid assembly with *SPAdes.* In a previous study, the use of *nanopolish* or *Racon* to polish prior to hybrid assembly with *SPAdes* resulted in accuracy exceeding 99.99% (98). Hybrid assembly with *SPAdes* has similar accuracy to *Canu* long read assemblies (98,99,100). In another study hybrid *SPAdes* was able to produce single circular plasmids, however in our study none of the isolates run with hybrid *SPAdes* resulted in complete plasmids carrying the ESBL gene (175). Instead a fragmentated assembly was obtained with the CTX-M gene located on one of the many divergent nodes of the chromosome where short read sequences failed to fully resolve the genome. Incorrect integration of plasmids into the bacterial chromosome has previously been reported with *SPAdes* (99).

The use of *plasmidSPAdes* was also attempted on a subset of samples because this was designed for plasmid assembly. However, this did not greatly improve plasmid completion, and resulted in fragmented assembly or incorporation into chromosomal

reads. This is in agreement with previous studies where *plasmidSPAdes* actually had less success in plasmid completion than hybrid *SPAdes* alone (14% and 22% completion respectively (100,176).

### 4.5.3 **Unicycler**

*Unicycler* first uses *SPAdes* to assemble Illumina reads followed by *Racon* in conjunction with *Miniasm* for hybrid assembly. *Racon* has been shown to improve median accuracy of reads from 89.8% to 99.25% (109). *Racon/Miniasm* (97.7%-99.32%) has similar accuracy to *Canu* (96.87%-98.61%), while also being an order of magnitude faster (109). *Unicycler* also uses *Pilon* for polishing, which has been shown to improve plasmid resolution in previous studies (100,101,139). The original paper describing *Unicycler* compared its performance to other hybrid assemblers. *Unicycler* assembled larger contigs with fewer misassemblies, even when long-read depth and accuracy were low (139). This is consistent with my study where *Unicycler* also performed better than *Miniasm/hybridSPAdes* and *Miniasm/plasmidSPAdes.*

### 4.5.4 ***Bowtie2*** **Filtration**

*Bowtie2* filtering was used to remove non-plasmid sequences prior to assembly in all cases where a complete plasmid was unable to be obtained with *Unicycler* alone. If a complete plasmid was obtained without *bowtie2* filtering, these samples were not rerun with *bowtie2* filtering; that is the best result was used for plasmid completion. Only a small subset of 5 (5/25, 20%) were rerun with the *bowtie2* prefilters for comparison and validation of the use of *bowtie2*. Most of these (4/5) gave similarly completed plasmids, with only one of these isolates resulting in a linear plasmid with the *bowtie2* prefilter. In addition, in a previous study, aggregation of the results of different methods (in that case *Canu*, *hybridSPAdes* and *plasmidSPAdes*) allowed the completion of more plasmids (29/37 vs 26/37) (100). A similar approach was taken in this study in either using *bowtie2* filtering or using *Unicycler* alone in order to maximise number of completely resolved plasmids as efficiently as possible. Since the same assembler is used

downstream of filtering, the only difference is in the short-read inputs to the assembler when filtration is applied. Removal of non-plasmid sequences may make the assembler less prone to error than when MinION plasmid reads are matched to Illumina whole genome sequences reads during local match assembly. In only 1 case (1/5, 20%) did *bowtie* filtering perform worse than without *bowtie* filtering. This may have been due to *bowtie2* removing reads in a non-uniform manner, especially in hard to assemble regions. In a previous study the use of *filtlong* to filter reads by quality and *Canu* for correction lead to inferior assemblies (177).

### 4.5.5 **Errors in Minion Results**

In the two cases of incongruence a plasmid was identified which carried a gene that was not identified in the *Nullarbor* report of the Illumina reads. One of the plasmids isolated was identified as a carrying a SHV-183 gene despite only CTX-M-14 and TEM-1B being identified by the Illumina sequence analysis using *Nullarbor*. In the other instance, a plasmid which carried CTX-M-15 and TEM-127 was identified from an isolate which only carried OXA9/SHV-12/TEM-1A β-lactamase genes according to the Illumina results. According to *AGICard* and the *EPI2me* resistance profile these incongruous genes were present in the raw MinION reads. This excludes the possibility of errors with *bowtie* or *Unicycler* further down the bioinformatic analysis pipeline. According to *CARD* and the *CGE ResFinder* the gene was absent in both the *bowtie2* filtered and unfiltered Illumina reads. Thus the anomalous genes were not missed by *Nullarbor* due to reference library differences. Furthermore, no β-lactamase gene was present in both sets of *bowtie2* filtered Illumina reads at all according to the CGE ResFinder. Thus *bowtie2* correctly filtered reads.

While, the MinION is known to have a high error rate, it is unlikely to explain the detection of a gene not present in the Illumina data (83,109). While isolates were subcultured and stored on several occasions since 2015, acquisition of a new gene is not plausible. Errors in Illumina sequencing are unlikely, as it is known to be more accurate than the MinION (177).

The most likely explanation is carryover contamination in the MinION reads. As this library preparation kit used contained 12 barcodes and these were reused between batches, it is possible that this may be due to contamination from previous runs despite flushing of the flow cell post use, as recommended. Though these barcodes were not used for isolates containing the respective incongruent gene in my experiments, the same flow cell had been used for other experiments with *E. coli* genes. Carryover contamination of barcodes from previous libraries run on the same flow cell despite an intervening wash step has been observed in other studies (94,178). These contaminant barcodes may also be of high quality, and therefore not excluded by quality cut-offs (94).

The use of a quality assessment tool such as *QUAST, Plantagora,* and *GAGE* to assess the quality of assemblies may have been useful. However, these tools have limitations: *QUAST* requires a known reference genome. We saw a wide variation in this set of isolates and natural variations may be mistaken for misassemblies in this case (179). In addition, the quality of the reads themselves have already been assessed.

## 4.6 IDENTIFICATION OF THE ESBL GENES

### 4.6.1 Dominance of CTX-M Genes

As expected given the rise to dominance of CTX-M genes worldwide, the proportion of CTX-M gene carrying *E. coli* (86.4%) far exceeded the proportion of isolates which only carried a TEM, SHV, or OXA gene (13.6%) according to Illumina WGS (5,9,15,18,19). Interestingly, this ratio is lower than in the 2016 ESR study, which describes the nationwide hospital and community ESBL populations in New Zealand over a 1 month period (August), where 99% of isolates carried CTX-M type and 1% carried a TEM, SHV, or OXA gene only (13). This difference may reflect Otago being a low prevalence area (17,21). Unfortunately, there was no ESR survey conducted in 2015, which is the year the samples in this study were collected; as of 2014, national surveys will be conducted every other year (13).

### 4.6.2 **Co-occurrence of CTX-M and TEM/OXA Genes**

The commonality of TEM and OXA genes co-occurring with CTX-M genes (27/66, 40.9%) is consistent with previous studies (180,181,182). These genes may be either located on the same plasmid, different plasmids, or on the chromosome. Of the isolates encoding both a CTX-M and a TEM or an OXA gene, the 14 of 16 (93.75%) plasmids isolated encoded only a CTX-M ESBL gene. There was only one plasmid isolated that encoded both CTX-M-15 and TEM-1B genes. There were also two plasmids isolated which encoded only a TEM gene- TEM-81 from a SHV-12/TEM-1B isolate and TEM-1B gene from an CTX-M-14/TEM-1B isolate. This suggests that the TEM or an OXA genes were likely encoded on different plasmids or on the chromosome. The TEM and OXA genes identified are known to have a broad spectrum β-lactam resistance phenotype and have been found in previous studies encoded on the chromosome and on plasmids (42,144-146,180-184).

Similarly, the CTX-M-14 gene of the isolate from which the TEM-1B carrying plasmid was isolated may have been encoded on different plasmids or the chromosome. Chromosomal integration of CTX-M type ESBL genes has been reported in bacterial isolates from both animal and human populations of Europe, Asia and less commonly in the Americas. CTX-M-14 integration in particular was found commonly in one study in Japan (9,185-190). Previous studies have also found TEM1B in conjunction with CTX-M type ESBL genes (CTX-M-1) on the same plasmid (42).

### 4.6.3 **TEM Genes on Plasmids**

Two plasmids isolated carried only a TEM β-lactamase gene. Plasmids encoding a TEM-1B (broad spectrum β-lactamase) gene and a TEM-81 (inhibitor resistant β-lactamase) gene were isolated from *E. coli* ST131 and ST636 isolates respectively. The plasmid with the broad-spectrum β-lactamase TEM-1B gene may also have encoded another

resistance mechanism (e.g. an efflux pump) which increased resistance to CTX during growth of the transformed isolates on CTX containing agar.

### 4.6.4 **SHV ESBL Genes**

No SHV genes were found in conjunction with CTX-M genes, though this co-occurrence has been described in the past (144-146,180-182). All SHV genes in this study were SHV-12, which is known to have an ESBL phenotype (144-146). SHV-12 encoding plasmids in *Enterobacteriaceae* have been found in aquatic environments in Switzerland, Portugal, Croatia, Canada, and China (59). They have also been found in food producing animals, companion animals, and wildlife in Spain, Portugal, Japan, Netherlands, Poland, the Czech Republic, China, Italy, USA, Mexico, and Germany. They have also been identified in vegetables in Switzerland, South Korea, Netherlands, and Spain (59). They have also been identified in human clinical specimens in Tunisia, France, Korea, Italy, France, the United Kingdom, Netherlands, Portugal, Spain, Taiwan, and Bulgaria (59).

### 4.6.5 **CTX-M ESBL Genes**

In the nationwide 2016 ESR survey, the proportions of CTX-M group 1 (CTX-M-15 and CTX-M-55) and group 9 (CTX-M-14 and CTX-M-27) genes in ESBL *E. coli* were approximately equal (50% and 49% respectively), though the proportion of group 1 genes was slightly higher than group 9 (55.8%/43.3%) in 2013 (13,29). In my study group 9 ESBL genes (53.1%) had a higher incidence than group 1 ESBL genes (30.3%). This is also in contrast to most regions worldwide where group 1 has been dominant, but is in line with China, South-East Asia, South Korea, Japan, Spain, and South America where group 9 CTX-M genes are the most common (15,19,22,23,25-27).

This difference is also reflected in the CTX-M types. In many countries, CTX-M-15 increased from the early 2000s (pre-2005) to the late 2000s (2005-2008); this has been reported in Spain, Portugal, Italy, Canada (which went from CTX-M-27 dominance to CTX-M-15 dominance), USA, Mexico, Bolivia, Israel, Egypt, China and Korea (though CTX-M-27 remained dominant in both), India, and South East Asia. While this increase continued in the 2009-2016 period in most countries, in other countries (Israel, Spain, North America (especially Canada), United Kingdom, Egypt, China, Vietnam, Turkey, France, Germany, Korea, and Japan) the proportion of CTX-M-15 dropped in favour of CTX-M-27, although not necessarily overtaking CTX-M-15 (22,34). There has also been a rise in CTX-M-1 and CTX-M-2 in some countries; not a single isolate of either was identified in my study (22).

Prior to the 2016 ESR survey, data of CTX-M types in New Zealand was not available; in previous surveys, genes were only typed to the group. Therefore, it is not possible to assess any trends in CTX-M proportions in New Zealand, though the proportion of group 9 ESBL *E. coli* has increased, suggesting the global trend may also apply to New Zealand. In Otago the incidence of CTX-M-27, CTX-M-14, and CTX-M-15 were not significantly different, though they represented the dominant genes over other CTX-M types, and the CTX-M-27 and CTX-M-14 proportions were higher than CTX-M-15, similar to the proportions reported from other countries, outlined above, where CTX-M-27 increased following 2009 (22). The 2016 ESR study only provided CTX-M gene proportions for all *Enterobacteriaceae* (including *Klebsiella* sp.) which were as follows: CTX-M-15, 54.5%; CTX-M-27, 22.7%; CTX-M-14, 14.2%; and CTX-M-55, 2.8%. Thus, the relative proportion of CTX-M-15 was lower in Otago in *E. coli* isolates.

Theories that have been put forward for the growing success of CTX-M-27, including their higher minimum inhibitory concentration (MIC) to ceftazidime compared to CTX-M-14 encoding isolates, a higher transmissibility in nosocomial settings, and transmission from animal sources where CTX-M-27 has been found commonly (22,31). The majority of the Otago isolates were from community setting (47/66, 71.2%) and the comparative MIC to ceftazidime was not assessed. There is no data available for whether the CTX-M-27 population has been increasing in Otago compared to the CTX-M-14

population, or the presence of CTX-M genes in animal populations in Otago or other parts of New Zealand.

Only four CTX-M- types were identified in this study. CTX-M types such as CTX-M-3 and CTX-M-1, CTX-M-24, CTX-M-9 and CTX-M-65 which were identified in the nationwide study were not identified in this study; this, along with the higher incidence of SHV ESBLs and OXA and TEM β-lactamases, may be attributed to the lower prevalence of ESBL *E. coli* in Otago (13,17,21). It may also represent under sampling due to the smaller number of isolates included in this study.

Of interest are the two isolates in this study which carried two CTX-M types, namely CTX-M-14 and CTX-M-55. This phenomena of co-existing ESBL genes has previously been observed. In a study in China, 4.2% of isolates carried two or more CTX-M types, with most of these (6/10, 60%) being CTX-M-14 and CTX-M-55, as found in the current study (27). The nationwide study in 2016 also identified one isolate containing both a group 1 and a group 9 CTX-M (information on CTX-M type not available) as well as a further isolate with CTX-M-64 (first identified in 2010 Japan), which is thought to be a chimeric hybrid of CTX-M-14 and CTX-M-15 or CTX-M-55 (13,32). The exact origin of CTX-M-64 is unclear due to disagreements between studies with small sample sizes, which presents a significant problem given the variety in genetic context observed around ESBL genes. In any case, homologous recombination to create these chimeric CTX-M genes would require the simultaneous presence of both CTX-M types in the same cell, a condition that is observed in the two CTX-M14 and CTX-M-55-encoding isolates from this study. We did not observe a plasmid carrying both CTX-M-14 and CTX-M-55 or a hybrid gene; we isolated plasmids that carried only CTX-M-14 from these two isolates.

## 4.7 PHYLOGENETIC ASSOCIATION WITH CTX-M GENES

MLST defines phylogenetic lineages in *E. coli* (39). Certain ESBL genes have been associated with certain sequence types, as defined by MLST, which accommodate transmission as they are highly virulent, multidrug resistant and highly transmissible (3,4,14-17,19). In contrast, a lot of diversity was observed in the CTX-M genes carried by isolates of many different MLST types in this set of isolates, with CTX-M-14, CTX-M-15 and CTX-M-55 carried on many different MLSTs, and no strong associations with any particular MLST. Only a limited number of significant associations between CTX-M genes and MLST types were observed.

### 4.7.1 ST131

An association between ST131 and CTX-M-15 has been described worldwide since the 2000s and has been proposed to be the main reason for the pandemic of ESBL *E. coli* in most countries worldwide (14,22,30,31,191,192). This association has been reported in the United Kingdom, Canada, Italy, Turkey, Croatia, USA, South Africa, Brazil, Norway, Spain, France, Portugal, Switzerland, Lebanon, India, Kuwait, South Korea, Central African Republic , Indonesia , Tunisia, Belgium, Denmark, Czech Republic, and Uruguay (14,22,30,31,191,192). Though the proportion of isolates with this association was more common in North America and Europe than in Asia (22). Additionally, in the 2016 ESR New Zealand study, an association between ST131 and CTX-M-15 was found (13).

The results of this study were in contrast to these national and international reports. While ST131 was the most common MLST which carried CTX-M-15 (7/15, 46.7%), ST131 was only significantly associated with CTX-M-27 (13/18, 72.2%, $p<0.01$). An association between CTX-M-27 and ST131 has been reported in other countries where CTX-M-27 incidence is increasing or is greater than CTX-M-15, and where there is a dominance of ST131. These countries include Israel, the Czech Republic, China, Switzerland, Japan, France, Portugal, Germany, and Vietnam; the increase in incidence of ST131 associated with CTX-M-27 was mostly seen in these countries after 2010 (or the

late 2000s) (22,30-37). Though in Japan CTX-M-27 was confined to ST131, this was not the case in Otago in this study, where more diversity was seen. Other countries where ST131 isolates carrying CTX-M-27 have been reported include Korea, Australia, Nepal, Cambodia, Spain, Netherlands, Canada, Denmark, and the USA (22,31,159,193). Overall, ST131 encoding CTX-M-27 may be considered an emerging issue.

Of interest is the presence of CTX-M-27-encoding ST131 *E. coli* isolates in the environment, in contrast to CTX-M-15-encoding ST131 isolates and non-ESBL ST131 isolates which are rarely found in the environment, including animals. CTX-M-27 producing ST131 has been identified in asymptomatic human faecal specimens in France, China, Portugal, and the Netherlands; samples from sick dogs and cats in Japan; samples from water birds from central Europe and Swiss rivers and lakes; and samples of well water from China (22,31). This suggests the possibility of an environmental reservoir for ST131/CTX-M-27 (22,31).

### 4.7.2 **ST38**

ST38 is not as ubiquitous as ST131, however it is one of the most common non-ST131 MLST types associated with CTX-M ESBL production. In this study ST38 isolates carried CTX-M-14, CTX-M-55, and CTX-M-27 genes. On phylogenetic analysis there was a small cluster of three ST38/CTX-M-27 isolates, one of only two clusters identified in this study. ST38 has most commonly been found worldwide in association with CTX-M-14. *E. coli* ST38 encoding CTX-M-14 in human infections has been identified in Mexico, China, Japan, Korea, Sweden, Norway, Spain, the Netherlands, and the United Kingdom (33). In CTX-M-14 producing *E. coli* isolates from urinary tract infections in a Korean study, ST38 was the second most commonly occurring sequence type (33). ST38 has previously been predominately associated with urinary tract infections (194). ESBL producing ST38 *E. coli* have also been obtained from other sources including water, pets and asymptomatic faecal carriage in humans in Switzerland, Germany, the Netherlands, Japan, and the United Kingdom (22,30,33,36,155,161,162,194).

## 4.8 NON-ESBL ANTIMICROBIAL DRUG RESISTANCE GENES

*Nullarbor* used *abricate* (v.0.4) and *resfinder* to screen the *SPAdes* assembly for antimicrobial resistance genes. Due to this, the resistance genes identified depend on the assembly generated and are limited by the library of *resfinder*. Partial hits were not considered for this analysis. The presence of a resistance gene does not necessarily result in phenotypic resistance to an antibiotic. ESBL *E. coli,* particularly ST131, frequently carry multiple resistance genes and has been associated with phenotypic MDR (3,4,10,15,16,38).

A *dfrA* gene was identified in most isolates in this study (41/66, 62.1%). This was consistent with the New Zealand nationwide 2016 ESR survey which identified a *dfrA* gene in 197/297 (66.3) *E. coli* isolates examined. A high incidence of *dfrA* genes in ESBL *E. coli* has also been reported internationally (14,159,195-198). The predominance of *dfrA17* over *dfrA1* seen in Otago (27/41, 65.9% and 6/41, 14.6% respectively), has also been seen internationally in Africa, Lebanon, Denmark, Netherlands, Korea, Australia and China (159,196,198). Though the opposite has been seen in other countries such as Spain, Portugal, and France (196). *dfrA12,* the other *dfrA* allele found in Otago has also been previously identified in China in ESBL *E. coli* isolated from humans and livestock (198,199).

The percentage of aminoglycoside modifying enzyme encoding genes identified in Otago (49/66, 74.2%) was also similar to the percentage found nationwide in the 2016 ESR survey 213/297 (71.7%). This high prevalence of aminoglycoside resistance genes (to gentamicin, streptomycin, spectinomycin, and kanamycin) has been observed in countries such as Canada, Australia, China (Hong Kong), the United Kingdom, India, as well as various African and European countries (3,4,14,17, 197,199-201). An association between gentamicin resistance genes and ST131 has been observed in Canada, Australia, Spain and the United Kingdom (3,4,14,17, 197,199-201). In our study 15/29 or 51.7% of isolates encoding resistance to gentamicin were ST131.

Plasmid mediated fluoroquinolone resistance genes occurred in 14/66 (21.2%) of isolates in Otago, and a similar incidence was seen in the ESR 2016 survey (70/297, 23.6%). Fluoroquinolone resistance associated with ESBL production has been reported internationally, in North America, Europe, Asia, Latin America, South Africa, the Middle East, and in the South Pacific, though incidence rates have varied from 6-79% (3,15,159,201-205). The incidence rates observed in Otago (21.2%) are similar to those reported overall in the South Pacific (23.5%) (202). The *qnrA* gene, identified in all continents except South America, has been associated with CTX-M-14; none of the isolates found in Otago expressed this combination (15,206). In Otago *qnrS* and *qnrB* alleles were observed, though with a very small incidence (4/66). In studies in Vietnam and France both *qnrS* and *qnrA* were identified, and in recent European surveys *qnrS* was found to be more common than *qnrA* (206). *qnrB* has been associated with CTX-M-15 or SHV-12 genes in international reports (15,207). In Otago both *qnrB* carrying isolates also carried SHV-12 genes. Combined resistance to aminoglycosides and quinolones in ESBL *E. coli* has been reported internationally (Europe, the Americas, and Asia [China and India]) (3,14,200,201,207). CTX-M-15 and ST131 have previously been associated with *aac(6')-Ib-cr*, which encodes resistance to quinolones as well as some aminoglycosides (tobramycin, gentamicin) in the USA, Europe, and South Africa (14,15,203,205,207,208). This was also observed in Otago with 8/10 (80%) isolates encoding *aac(6')-Ib-cr* also encoding CTX-M-15 (the other two encoding CTX-M-27) and 7/10 (70%) were ST131.

High rates of tetracycline resistance genes were observed in Otago (39/66, 59.1%), similarly to international reports in Asia and Africa (198,199).

We identified 3 isolates carrying *fosA* genes which encode resistance to fosfomycin; no genes of this type were identified in *E. coli* in the ESR survey in 2016. This resistance gene has however been reported worldwide (Europe, Asia, South America, and the Middle East) in ESBL *E. coli* isolates with incidence ranging from 0-19% in humans and livestock (205,209,210). Colistin resistance genes, which had been identified in ESBL *E. coli* human and environmental sources in other countries (albeit in small numbers),

were not observed both in Otago in my study and nationwide in the ESR survey, although there was a portion of a mcr-1 gene in one isolate (211).

MDR genes have previously been reported to occur on the same plasmid as CTX-M genes, constituting "resistance islands" (38,39). This was not assessed in this study but may be interesting for future studies, which can use the hybrid plasmid assemblies produced by this study to delineate the genetic context of non-ESBL resistance genes.

### 4.8.1 **Comparison with Phenotype**

In the NZ ESR survey chromosomally mediated fluoroquinolone resistance (chromosomal mutations in *gyrA* and *parC*) occurred in a majority of the isolates (181/297 (60.9%)) while plasmid mediated fluoroquinolone resistance occurred in only 70/297 (23.6) of isolates (13). Chromosomally mediated quinolone resistance was not assessed in my study. This would likely account for the high percentage (38/51, 74.5%) of isolates which showed phenotypic resistance to ciprofloxacin but had no detectable quinolone (plasmid mediated) resistance gene. This phenomenon of chromosomal quinolone resistance in ESBL *E. coli* has also been observed overseas (India, USA, Netherlands) in human, animal and environmental specimens (208,212). In my study a *dfrA* gene was identified in 90% of the isolates reported to be trimethoprim resistant. The 2016 New Zealand ESR survey had a similar high rate of congruence between genotype and phenotype at 87.9% (13). Future studies on the phenotypic resistance to tetracyclines and fosfomycin for this set of isolates would be useful to determine the association between phenotypic expression of resistance and resistance genes.

## 4.9 ESBL-ENCODING PLASMID TYPES

Plasmid transmission has been important in the spread of ESBL *E. coli* (18). ESBL-encoding plasmids were isolated from 41 of 66 (62.1%) ESBL *E. coli*. The most common plasmid types were IncF plasmids (22/41, 53.7%). IncF plasmids are diverse, with multiple replicon types such as FIA, FIB, and FII. In this study most of the ESBL carrying IncF plasmids were IncFIA/IncFII (9/22, 40.9%) and IncFII (10/22, 45.5%) followed by IncFIA (2/22, 9.1%) and IncFIB/IncFII (1/22, 4.5%).

IncF plasmids have a narrow host-range but are well adapted to *Enterobacteriaceae* and show high conjugative ability and high transmission rates among *Enterobacteriaceae* (4,14,15,19,50,51,60). IncF plasmids also have a propensity to acquire multiple resistance genes, which allows co-selection (4,14,15,19,50,51,60). IncF plasmids (including IncFIB, IncFIA, IncFI and IncFII) are the most common ESBL carrying plasmids across all CTX-M-groups worldwide (the Americas, Europe, Africa, and Asia) (14,19,35,50,51,166,191,197,213-218). The results of my study are consistent with these international reports, with 65.6% (21/32) of CTX-M encoding plasmids typing as IncF. Internationally (North America, Africa, and Europe [Germany, Norway, Czech Republic]), a strong association between CTX-M-15 and IncF plasmids (namely IncFII and IncFIA/IncFIB multi-replicon plasmids) has been reported (14,19,66,191,197,212,218-221). In contrast, in this study CTX-M-15 was rarely associated with IncF plasmids (1/9, 11.1%). Instead, IncF plasmids most commonly carried CTX-M-14 and CTX-M-27 genes.

Almost all CTX-M-27 genes were carried on IncF plasmids; only 1 of 16 (6.25%) CTX-M-27-encoding plasmids was non-IncF (IncB/O/K/Z). Specifically, these IncF plasmids were IncFII, IncFIA, and IncFIA/IncFII plasmids. An association of CTX-M-27 with IncF plasmids has been reported in other countries. A study in Japan found an association between CTX-M-27 and IncF plasmid types IncFII, IncFIA, and IncFIB (155). In a study in China CTX-M-27 was associated with FII/FIA/FIB in *E. coli* from animal sources (213). In Vietnam CTX-M-27 was found in plasmid types IncFI, IncFIA, IncFIA/IncFII,

IncFII/IncFIB/IncL1 (listed in descending order of incidence) in *E. coli* from both human and animal sources (35). In Germany CTX-M-27 was associated with IncFIA/ IncFIB plasmids (34).

A greater variety of plasmids carried CTX-M-14, with 3 of 8 (37.5%) being non-IncF plasmids. CTX-M-14 encoding plasmids were IncFII, Inc FIA, IncFIA/IncFII, IncB/O/K/Z, and IncI1. In international reports CTX-M-14 has also been found to be carried on a variety of plasmid types, including on IncF (Hong Kong, Korea and France), IncK (Spain and the United Kingdom) and IncI1 (Uruguay) (19,22,50,51,192,214,217).

In this study, CTX-M-15 was carried mostly on IncI1 plasmids (4 of 9, 44.4%). CTX-M-15 carrying IncI1 plasmids have previously been described in the USA, Vietnam, China, and Germany (14,35,66,212,214,222,223). IncI1 was the third most common plasmid type found in Otago in this study and was also found carrying CTX-M-55, CTX-M-14, and TEM/SHV. IncI1 plasmids are of particular interest as they are highly transmissible; they are globally spread among *E. coli* populations from both humans and animals and appear to be widely distributed among different members of the *Enterobacteriaceae* family (42). However, it should be noted that in this study, 3 isolates where the plasmid incompatibility type was not able to be determined carried CTX-M-15 genes. CTX-M-15 was also carried on IncFIA/IncFII and Col156 plasmids. CTX-M-15 has been found on IncHI2, IncK, IncX, IncN, IncA/C, and ColBS512 in other studies, if less commonly than on IncF plasmids (14,35,212,214,216,222).

The higher incidence of IncF plasmids encoding CTX-M-27 compared to IncF plasmids encoding CTX-M-15 parallels the association of ST131 with CTX-M-27 rather than CTX-M-15 in this study, as discussed above. A strong relationship between ST131 and IncF epidemic plasmids (especially FIA and FII replicon types) has been seen internationally (34,216,224). This is also seen in this set of isolates, with most (12/17) ST131 carrying ESBL genes on IncF plasmids, albeit not encoding CTX-M-15 (224).

Multireplicon plasmids were common in our set of isolates. In fact, the most commonly

found plasmid type was IncFIA/IncFII multireplicon plasmids (equal in number to IncFII). We also described a single IncFIB/IncFII and a single IncN/IncX4 plasmid. Multireplicon plasmids harbouring IncFII in addition to IncFIA and/or IncFIB have previously been described extensively as well as IncFIA/IncFIB plasmids (19,53,216,217,224). Conversely, there have been reports from Turkey, Tanzania, Germany, Egypt, and India, of high incidence of IncFIA and IncFIB plasmids (alone and in combination), with low rates of IncFII, carrying different CTX-M types (53). IncFII plasmids are of interest having been identified in environmental and animal sources (225).

An IncN/X4 plasmid encoded CTX-M-55. IncN is important as a broad host-range plasmid type, with high conjugation frequency, and has been found in the bacterial flora of humans, animals and the environment, making it conducive to the spread of ESBL genes (58,217). It has also been associated with MDR (58,217). CTX-M-55 was also carried on IncI1 and IncX1 plasmids in this study. CTX-M-55 has previously been associated with IncFII, IncFIA, IncHI2, and IncI1 (35,213,220).

SHV-type ESBLs have also been identified on a variety of plasmids, from narrow range (IncF, IncI1, and IncX3), to broad range (IncA/C, IncHI2) (59). In this study SHV plasmids were all narrow range; being mostly IncX3 (four isolates) and one IncFII/FIB. IncX is important as it is associated with MDR (217).

A further two plasmids, which encoded a broad spectrum β-lactamase gene (TEM-1B) and an inhibitor resistant β- lactamase gene (TEM-81), were carried on IncFII/FIA and an untypeable plasmid respectively. This is in contrast to previous studies where TEM-type beta-lactamase genes have been associated with IncA/C and IncI1-type plasmids (42,60).

Four plasmids were not able to be typed in this study. One of these was not able to be completed using the plasmid assembly pipeline. This may be due to insufficient quality

of the MinION reads, resulting in an incomplete assembly which lacked the sequence regions used for plasmid typing. The inability to type the other three completed plasmids may be due to the natural mutations and rearrangements of plasmids occurring in the sequence regions used for plasmid typing. Many other studies have reported plasmids which were designated untypeable (217,226).

## 4.10 INSERTION SEQUENCES ASSOCIATIONS

Mobile genetic elements can mediate the movement of resistance genes between plasmids (2,14,38). In this study the insertion sequences surrounding the ESBL genes were identified. Overall, significant diversity was seen. There were 13 different insertion sequence combinations upstream and 12 downstream. Each ESBL gene had many different surrounding insertion sequences, although some common insertion sequences as well as associations between ESBL genes and insertion sequences were seen.

There were differences seen in the genetic context of different groups of ESBL genes. The association of downstream orf*477* (with or without an intact or truncated Tn*3* transposon downstream) with group 1 CTX-M genes has been observed in other studies in isolates from humans and animals (14,66,216,227,228). The occurrence of IS*26* downstream of orf*477*, seen in this study, has also been previously observed in a study conducted in Spain (229). The exclusive association of Group 9 CTX-M genes (CTX-M-14 and CTX-M-27) with IS*903D* downstream has also previously been reported globally (25,66,191,216,229,230). These reports also found a varying presence of IS*26* downstream of IS*903D*, and a varying presence of *iroN*-Tn*3*-IS*26* downstream of IS*903D* (25,66,191,216,229,230). TEM-1B flanked by *tnpR*, *tnpA*, Tn*2* and IS*26* has been previously observed (42). IS*3000* has previously been observed in the genetic context of group 9 CTX-M genes, however in this study IS*3000* only appeared upstream of SHV ESBL genes (19).

IS*Ecp1* and IS*26* insertion sequences were commonly seen in the genetic context of all CTX-M genes seen in this study. IS*Ecp1* and IS*26* are well documented as the most

common elements associated with CTX-M-genes (19,22,66,191,215,216,229,231-233). IS*Ecp1* occurred upstream (5') of all CTX-M genes, while the presence of IS*26* varied. IS*26* occurred less commonly upstream of CTX-M-14 (25%, 2/8) and CTX-M-15 (22%, 2/9) than of CTX-M-27 (94%, 15/16) and CTX-M-55 (50%, 2/4). The IS*26*-IS*Ecp1* element occurred in 94% of CTX-M-27 isolates (15/16), in contrast 2/4 CTX-M-55 isolates (50%) had this element, with only one occurrence in CTX-M-14 isolates (1/8, 12.5%) and CTX-M-15 isolates (1/8, 12.5%). This pattern has been observed in a previous large-scale study in Japan, though the study did not assess for the presence of Tn*A*, which occurred upstream of the IS*26*-IS*Ecp1* element in most CTX-M-27 isolates (14/16, 87.5%) in Otago (25). The IS*26*-IS*Ecp1* element has also previously been found to be associated with ST131 and IncF plasmids (216,229). However, in this study only the CTX-M-27 ST131 isolates (10/11, 90.9%) and 1 CTX-M-55 ST131 (1/1) isolate had this element (11/17, 64.7% of ST131 in total). None of the ST131 isolates carrying CTX-M-14 or CTX-M-15 genes contained this structure. In addition, IS*26*-IS*Ecp1* was also present in all other STs which carried CTX-M-27. Nevertheless, the frequency of IS*26* presence suggests IS*26*-mediated transposition of CTX-M genes (216). IS*26* also commonly appeared downstream (3') of the CTX-M gene (23/43, 53.5%), whereas IS*Ecp1* was only found upstream, consistent with other studies (25,216,229).

In comparison to CTX-M-14, which mostly had only IS*Ecp1* upstream (6/8, 75%), CTX-M-15 also had a Tn*A* sequence upstream of IS*Ecp1* (4/8, 50%); this difference has previously been observed (216). The varying occurrence of Tn*3* structures upstream or downstream of ESBL genes in other instances in this study has also been observed in other studies (215,216). Tn*3*-type transposons are thought to be a target entry point for integration of ESBL genes and surrounding insertion sequences into the plasmid (216). IS*CR1* is an insertion sequence that has been important in the dissemination of CTX-M genes but was not identified in any isolates in this study (19,215).

There were differences downstream of the group 9 CTX-M isolates as well. With IS*903D*-*iroN*-Tn*3*-like-IS*26*-Tn*A* (designated S) being found only in CTX-M-27 isolates (4/16, 25%). However, the same structure without the downstream Tn*A* (designated T) was

found in both CTX-M-14 (2/8, 25%) and CTX-M-27 (2/16 12.5%). This downstream structure (IS*903D*-*iroN*-Tn*3*-like-IS*26*) has been previously described in Japan also with low incidence, but only in the context of CTX-M-27 (25). In summary, the Tn*A*-IS*26*-IS*Ecp1*-CTX-M-27-IS*903D*-IS*26* unit (E-V) which was the most common genetic context of CTX-M-27 has been observed in previous studies, though some studies did not assess the presence of Tn*A* as discussed previously (25,216). An additional isolate with E-CTX-M-27-V had bilateral presence of IS*1b;* to my knowledge, this has not been seen in the context of CTX-M-27 in previous published studies. IS*Ecp1*-CTX-M-14-IS*903D* unit (A-W) was the most common in CTX-M-14.

In group 1 isolates, all four CTX-M-55 isolates had different insertion sequences. Though 3/4 (75%) CTX-M-55 had orf*477* only downstream of the CTX-M gene, with only one orf*477*-Tn*A* occurrence, upstream insertion sequences differed. In contrast in CTX-M-15, the orf*477* gene only occurred in combination with either Tn*A* (5/8, 62.5%) or Tn3-IS26 (3/8, 37.5%) unit. In summary, only Tn*A*-IS*Ecp1*-CTX-M-15-orf*477*-Tn*A* (B-Y) (3/8,37.5%) and IS*Ecp1*-CTX-M-15-orf*477*-Tn*A* (A-Y) (2/8, 25%) occurred in more than one isolate among CTX-M-15 isolates. The IS*Ecp1*-CTX-M-15-orf*477* and IS*Ecp1*-CTX-M-55-orf*477* units have been described in previous studies (25,216,234,235).

All four SHV-12 isolates had different insertion sequences. These included an isolate with only IS*3000*, IS*Ehe3,* and IS*26* upstream and no identified downstream insertion sequences, an IS*26*-SHV-12-*deoR* isolate (L-P), a Tn*A*-IS*26*-SHV-12-*deoR* (I-P) isolate, and an IS*26*-SHV-12-Tn*A* (L-O) isolate. IS*26*-SHV-12-*deoR* units have been identified in past studies in both ESBL *E. coli*, other *Enterobacteriaceae* and *P. aeruginosa* (59,236-238).

## 4.11 OVERALL GENETIC DIVERSITY

While there were some observed associations between ESBL genes, their genetic context (including plasmid type), *E. coli* MLST type, and clusters on the pangenome, as discussed above, overall considerable genetic diversity was seen. All the TEM genes were different in terms of MLST, plasmid type, and genetic context. The four CTX-M-55 isolates were all different MLSTs, had different genetic contexts, and were carried on plasmids of three different incompatibility types. While there were three CTX-M-15/ST131 isolates, they differed in their genetic contexts. The ESBL encoding plasmids of the three CTX-M-14/ST405 isolates were of different incapability types and had differing insertion sequences around the CTX-M-14 gene. Furthermore, where the plasmid type, genetic context and ESBL gene were the same, the MLST type of the isolate varied. For example, the three IncI1/B-Y/CTX-M-15 plasmids were from ST131, ST162, and ST3036 isolates. In another instance the two IncB/O/K/Z/A-W/CTX-M-14 plasmids were from ST 69 and ST405 isolates. Among the four SHV-12 isolates, there were three isolates which were ST636, carried the SHV gene on plasmids of incompatibility type IncX3, however the insertion sequences differed (L-O, L-Q, and H-). In the CTX-M-27 isolates there were some notable clones. Clusters of two (2/66, 3.0%) CTX-M-27/ST131/IncFII/ET isolates, three (3/66, 4.5%) ST38/IncFII/ES isolates, and six (6/66, 9.1%) ST131/IncFIA-IncFII/EV isolates were identified. The other 5 CTX-M-27 isolates differed in MLST, plasmid incompatibility type and/or insertion sequences.

This observed low level clustering, with the overall population of ESBL *E. coli* strains encoding multiple enzyme types, sequence types, genetic contexts, and plasmid types, indicates multiple introductions into the Otago population with limited evidence of horizontal transmission (39,74). This is in contrast to a limited number of point source introductions and onward transmission of highly transmissible drug resistant strains that would have resulted in a few dominant strains of the same MLST profile, carrying the same ESBL gene, in the same genetic context on the same plasmid type (27,39,71-74).

These multiple introductions could be from numerous sources. Inadequate,

unnecessary, or overuse of antibiotics in treatment, overuse of antibiotics in agriculture could all constitute selection pressures for the mobilization of antibiotic resistance genes (68,239,240,241,242). Recent studies have shown that the low antibiotic concentrations that have been found in environmental residues (such as hospital and wastewater effluent) are sufficient for the positive selection of ESBL resistance (i.e. higher than the minimum selective concentration) (68,240,242). Many of the ESBL and associated sequence types and plasmids types found in this study have been found commonly from animal and environmental sources in international studies as discussed previously. Future studies on the presence of ESBL genes in the environment, animals, and asymptomatic faecal carriage in humans could be important for the elucidation of potential reservoirs. The finding that the increase in ESBL urinary tract infection incidence in Otago is largely not due to the transmission of highly virulent clones is also important to avoid inappropriate outbreak management (242).

Duplicate isolates, that is isolates from the same patient from different periods of time, were excluded as they had not been cryopreserved, so determination of mutation rate was not possible. This would have allowed us to observe any change over time in the infecting strain, and the rate of change in the genome, in order to accurately estimate the molecular clock for *E. coli* in vivo (74).

Molecular evidence of clonal linkage merely provides support to epidemiological evidence (39). For example, inappropriate antibiotic use in medicine or agriculture exerts a selective pressure, which may result in the emergence and spread of antibiotic resistance (4,15,38,43). Further epidemiological evaluations, which were not carried out in this study, could help identify person-to-person contact or common exposure to epidemiological sources. Although, the ESR nationwide study in 2016 on ESBL *Enterobacteriaceae* found no significant association between patient demographics and phylogeny or ESBL genes, this was very high level and lacks granularity compared to contact network epidemiology and molecular studies (13).

# 5 CONCLUSION

This study provides up to date information on the ESBL *E. coli* causing urinary tract infections in Dunedin and the genetic context of resistance. The development of a bioinformatic pipeline for the hybrid assembly of Illumina short read WGS and MinION long read sequences of plasmids allowed the full genetic characterisation of isolates easily and quickly. Taking into account ESBL gene, sequence type, plasmid type, and insertion sequences all fully characterised isolates showed high diversity, with only 2 small clusters of clonal isolates. Instead of the expansion of one or a few bacterial clones, and transmission of mobile genetic element(s) encoding multiple resistance genes between strains, there appear to have been multiple introductions of ESBL genes into Otago, a low prevalence area. This raises important questions about the reservoir of ESBL genes and the vehicle(s) of transmission. Further combined laboratory and epidemiological studies will be required to investigate this. This data can inform the need for infection control and/or antibiotic policy assessments in Otago and potentially other low prevalence areas.

# 6 REFERENCES


1. WHO (2014) Antimicrobial Resistance: Global Report on Surveillance.: WHO Press, 2014.

2. Doumith M, Day M, Ciesielczuk H, Hope R, Underwood A, Reynolds R, Wain J, Livermore DM, Woodford N. Rapid identification of major *Escherichia coli* sequence types causing urinary tract and bloodstream infections. Journal of Clinical Microbiology. 2015 Jan 1;53(1):160-6.

3. Villa L, García-Fernández A, Fortini D, Carattoli A. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. Journal of Antimicrobial Chemotherapy. 2010 Oct 8;65(12):2518-29.

4. Brolund A, Sandegren L. Characterization of ESBL disseminating plasmids. Infectious Diseases. 2016 Jan 2;48(1):18-25.

5. Novais Â, Comas I, Baquero F, Cantón R, Coque TM, Moya A, González-Candelas F, Galán JC. Evolutionary trajectories of beta-lactamase CTX-M-1 cluster enzymes: predicting antibiotic resistance. PLoS Pathogens. 2010 Jan 22;6(1)

6. O'Neill J. Tackling drug-resistant infections globally: Final report and recommendations. 2016. HM Government and Welcome Trust: UK. 2018.

7. Conly JM, Johnston BL. Where are all the new antibiotics? The new antibiotic paradox. Canadian Journal of Infectious Diseases and Medical Microbiology. 2005;16(3):159-60.

8. Elander RP. Industrial production of β-lactam antibiotics. Applied Microbiology and Biotechnology. 2003 Jun 1;61(5-6):385-92.

9. Bonnet R. Growing group of extended-spectrum β-lactamases: the CTX-M enzymes. Antimicrobial Agents and Chemotherapy. 2004 Jan 1;48(1):1-4.

10. Paterson DL, Bonomo RA. Extended-spectrum β-lactamases: a clinical update. Clinical Microbiology Reviews. 2005 Oct 1;18(4):657-86.

11. Foxman B. The epidemiology of urinary tract infection. Nature Reviews Urology. 2010 Dec;7(12):653.

12. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nature Reviews Microbiology. 2004 Feb;2(2):123.

13. Heffernan H, Woodhouse R, Draper J, Ren X, 2016 survey of extended-spectrum β-lactamase-producing *Enterobacteriaceae*. Antimicrobial Reference Laboratory and Helath Group, Institute of Environmental Science and Research Limited (ESR); July 2018.

14. Mathers AJ, Peirano G, Pitout JD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant *Enterobacteriaceae.* Clinical Microbiology Reviews. 2015 Jul 1;28(3):565-91.

15. Cantón R, Coque TM. The CTX-M β-lactamase pandemic. Current Opinion in Microbiology. 2006 Oct 31;9(5):466-75.

16. Nicolas-Chanoine MH, Bertrand X, Madec JY. *Escherichia coli* ST131, an intriguing clonal group. Clinical Microbiology Reviews. 2014 Jul 1;27(3):543-74.

17. Petty NK, Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD, Moriel DG, Peters KM, Davies M, Rogers BA. Global dissemination of a multidrug resistant *Escherichia coli* clone. Proceedings of the National Academy of Sciences. 2014 Apr 15;111(15):5694-9.

18. D'Andrea MM, Arena F, Pallecchi L, Rossolini GM. CTX-M-type β-lactamases: a successful story of antibiotic resistance. International Journal of Medical Microbiology. 2013 Aug 1;303(6-7):305-17.

19. Naseer U, Sundsfjord A. The CTX-M conundrum: dissemination of plasmids and *Escherichia coli* clones. Microbial Drug Resistance. 2011 Mar 1;17(1):83-97.

20. Woerther PL, Burdet C, Chachaty E, Andremont A. Trends in human fecal carriage of extended-spectrum β-lactamases in the community: toward the globalization of CTX-M. Clinical Microbiology Reviews. 2013 Oct 1;26(4):744-58.

21. Williamson DA, Heffernan H. The changing landscape of antimicrobial resistance in New Zealand. The New Zealand Medical Journal (Online). 2014 Sep 26;127(1403):42.

22. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. Journal of Antimicrobial Chemotherapy. 2017 May 25;72(8):2145-55.

23. Rocha FR, Pinto VP, Barbosa FC. The spread of CTX-M-Type Extended-Spectrum β-Lactamases in Brazil: a systematic review. Microbial Drug Resistance. 2016 Jun 1;22(4):301-11.

24. Barlam TF, Gupta K. Antibiotic resistance spreads internationally across borders. The Journal of Law, Medicine & Ethics. 2015 Jun 1;43(S3):12-6.

25. Matsumura Y, Johnson JR, Yamamoto M, Nagao M, Tanaka M, Takakura S, Ichiyama S, Fujita N, Komori T, Yamada Y, Shimizu T. CTX-M-27-and CTX-M-14-producing, ciprofloxacin-resistant *Escherichia coli* of the H30 subclonal group within ST131 drive a Japanese regional ESBL epidemic. Journal of Antimicrobial Chemotherapy. 2015 Jun 1;70(6):1639-49.

26. Xia S, Fan X, Huang Z, Xia L, Xiao M, Chen R, Xu Y, Zhuo C. Dominance of CTX-M-type extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* isolated from patients with community-onset and hospital-onset infection in China. PLoS One. 2014 Jul 1;9(7).

27. Zhang J, Zheng B, Zhao L, Wei Z, Ji J, Li L, Xiao Y. Nationwide high prevalence of CTX-M and an increase of CTX-M-55 in *Escherichia coli* isolated from patients with community-onset infections in Chinese county hospitals. BMC Infectious Diseases. 2014 Dec 3;14(1):1.

28. Rogers BA, Ingram PR, Runnegar N, Pitman MC, Freeman JT, Athan E, Havers SM, Sidjabat HE, Jones M, Gunning E, De Almeida M. Community-onset *Escherichia coli* infection resistant to expanded-spectrum cephalosporins in low-prevalence countries. Antimicrobial agents and Chemotherapy. 2014 Apr 1;58(4):2126-34.

29. Heffernan H, Woodhouse R. Annual Survey of extended-spectrum β-lactamase (ESBL)-producing *Enterobacteriaceae*, 2014. Institute of Environmental Science & Research Limited. 2014.

30. Cantón R, González-Alba JM, Galán JC. CTX-M enzymes: origin and diffusion. Frontiers in Microbiology. 2012 Apr 2;3:110.

31. Matsumura Y, Pitout JD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, DeVinney R, Bradford PA, Motyl MR, Tanaka M. Global *Escherichia coli* sequence type 131 clade with blaCTX-M-27 gene. Emerging Infectious Diseases. 2016 Nov;22(11):1900.

32. He D, Partridge SR, Shen J, Zeng Z, Liu L, Rao L, Lv L, Liu JH. CTX-M-123, a novel hybrid of the CTX-M-1 and CTX-M-9 group β-lactamases recovered from *Escherichia coli* isolates in China. Antimicrobial Agents and Chemotherapy. 2013 Aug 1;57(8):4068-71.

33. Merida-Vieyra J, De Colsa A, Castañeda YC, Barbosa PA, Andrade AA. First Report of Group CTX-M-9 Extended Spectrum Beta-Lactamases in *Escherichia coli* Isolates from Pediatric Patients in Mexico. PloS one. 2016 Dec 19;11(12)

34. Ghosh H, Doijad S, Falgenhauer L, Fritzenwanker M, Imirzalioglu C, Chakraborty T. blaCTX-M-27–Encoding *Escherichia coli* Sequence Type 131 Lineage C1-M27 Clone in Clinical Isolates, Germany. Emerging infectious diseases. 2017 Oct;23(10):1754.

35. Nguyen VT, Jamrozy D, Matamoros S, Carrique-Mas JJ, Ho HM, Thai QH, Nguyen TN, Wagenaar JA, Thwaites G, Parkhill J, Schultsz C. Limited contribution of non-intensive chicken farming to ESBL-producing *Escherichia coli* colonization in humans in Vietnam: an epidemiological and genomic analysis. Journal of Antimicrobial Chemotherapy. 2019 Mar 1;74(3):561-70.

36. Gerhold G, Schulze MH, Gross U, Bohne W. Multilocus sequence typing and CTX-M characterization of ESBL-producing *E. coli*: a prospective single-centre study in Lower Saxony, Germany. Epidemiology & Infection. 2016 Nov;144(15):3300-4.

37. Birgy A, Bidet P, Levy C, Sobral E, Cohen R, & Bonacorsi S. CTX-M-27–Producing *Escherichia coli* of Sequence Type 131 and Clade C1-M27, France. Emerging Infectious Diseases. 2017 May;23(5): 885.

38. Partridge SR. Resistance mechanisms in *Enterobacteriaceae*. Pathology-Journal of the RCPA. 2015 Apr 1;47(3):276-84.

39. Filippis I, McKee ML, editors. Molecular typing in bacterial infections. Springer Science & Business Media; 2012 Nov 7.

40. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. Clinical Microbiology Reviews. 2018 Oct 1;31(4):e00088-17. Figure 1,Examples of mobile genetic elements (MGE) and processes involved in intracellular mobility or intercellular transfer of antibiotic resistance genes;p.3.

41. Stokes HW, Gillings MR. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. FEMS Microbiology Reviews. 2011 Sep 1;35(5):790-819.

42. Wang J, Stephan R, Zurfluh K, Hächler H, Fanning S. Characterization of the genetic environment of blaESBL genes, integrons and toxin-antitoxin systems identified on large transferrable plasmids in multi-drug resistant *Escherichia coli.* Frontiers in Microbiology. 2015 Jan 6;5:716.

43. Partridge SR. Analysis of antibiotic resistance regions in Gram-negative bacteria. FEMS Microbiology Reviews. 2011 Sep 1;35(5):820-55.

44. Poirel L, Bonnin RA, Nordmann P. Genetic support and diversity of acquired extended-spectrum β-lactamases in Gram-negative rods. Infection, Genetics and Evolution. 2012 Jul 31;12(5):883-93.

45. Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, Nolan LK. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. Applied and Environmental Microbiology. 2007 Mar 15;73(6):1976-83.

46. Novick RP. Plasmid incompatibility. Microbiological reviews. 1987 Dec;51(4):381.

47. Couturier MA, Bex F, Bergquist PL, Maas WK. Identification and classification of bacterial plasmids. Microbiological Reviews. 1988 Sep;52(3):375.

48. Ying J, Wu S, Zhang K, Wang Z, Zhu W, Zhu M, Zhang Y, Cheng C, Wang H, Tou H, Zhu C. Comparative genomics analysis of pKF3-94 in *Klebsiella pneumoniae* reveals plasmid compatibility and horizontal gene transfer. Frontiers in Microbiology. 2015 Aug 18;6:831.

49. Villa L, García-Fernández A, Fortini D, Carattoli A. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. Journal of Antimicrobial Chemotherapy. 2010 Dec 1;65(12):2518-29.

50. Carattoli A. Resistance plasmid families in *Enterobacteriaceae*. Antimicrobial Agents and Chemotherapy. 2009 Jun 1;53(6):2227-38.

51. Carattoli A. Plasmids and the spread of resistance. International Journal of Medical Microbiology. 2013 Aug 31;303(6):298-304.

52. Ahmed MY. Characterization and molecular epidemiology of Extended-Spectrum-ß-Lactamase-Producing *Escherichia coli* derived from University Hospitals of Egypt and Germany. [unpublished thesis] Sohaj, Egypt: Institute of Medical Microbiology: Justus-Liebig-University Giessen; 2013.

53. Ali SZ, Ali SM, Khan AU. Prevalence of IncI1-Iγ and IncFIA-FIB type plasmids in extended-spectrum β-lactamase-producing *Klebsiella pneumoniae* strains isolated from the NICU of a North Indian hospital. Microbiology. 2014 Jun 1;160(6):1153-61.

54. Mshana SE, Imirzalioglu C, Hossain H, Hain T, Domann E, Chakraborty T. Conjugative IncFI plasmids carrying CTX-M-15 among *Escherichia coli* ESBL producing isolates at a University hospital in Germany. BMC Infectious Diseases. 2009 Jun 17;9(1):1.

55. Mshana SE, Fritzenwanker M, Falgenhauer L, Domann E, Hain T, Chakraborty T, Imirzalioglu C. Molecular epidemiology and characterization of an outbreak causing *Klebsiella pneumoniae* clone carrying chromosomally located bla CTX-M-15 at a German University-Hospital. BMC Microbiology. 2015 Jun 17;15(1):1.

56. Ewers C, Bethe A, Stamm I, Grobbel M, Kopp PA, Guerra B, Stubbe M, Doi Y, Zong Z, Kola A, Schaufler K. CTX-M-15-D-ST648 *Escherichia coli* from companion animals and horses: another pandemic clone combining multiresistance and extraintestinal virulence?. Journal of Antimicrobial Chemotherapy. 2014 May 1;69(5):1224-30.

57. Gonullu N, Aktas Z, Kayacan CB, Salcioglu M, Carattoli A, Yong DE, Walsh TR. Dissemination of CTX-M-15 β-lactamase genes carried on Inc FI and FII plasmids among clinical isolates of *Escherichia coli* in a university hospital in Istanbul, Turkey. Journal of Clinical Microbiology. 2008 Mar 1;46(3):1110-2.

58. Dolejska M, Villa L, Hasman H, Hansen L, Carattoli A. Characterization of IncN plasmids carrying blaCTX-M-1 and qnr genes in *Escherichia coli* and *Salmonella* from animals, the environment and humans. Journal of Antimicrobial Chemotherapy. 2013 Feb 1;68(2):333-9.

59. Liakopoulos A, Mevius D, Ceccarelli D. A review of SHV extended-spectrum β-lactamases: neglected yet ubiquitous. Frontiers in Microbiology. 2016 Sep 5;7:1374.

60. Marcadé G, Deschamps C, Boyd A, Gautier V, Picard B, Branger C, Denamur E, Arlet G. Replicon typing of plasmids in *Escherichia coli* producing extended-spectrum β-lactamases. Journal of Antimicrobial Chemotherapy. 2009 Jan 1;63(1):67-71.

61. Smet A, Van Nieuwerburgh F, Vandekerckhove TT, Martel A, Deforce D, Butaye P, Haesebrouck F. Complete nucleotide sequence of CTX-M-15-plasmids from clinical *Escherichia coli* isolates: insertional events of transposons and insertion sequences. PLoS One. 2010 Jun 18;5(6):e11202.

62. Karim A, Poirel L, Nagarajan S, Nordmann P. Plasmid-mediated extended-spectrum β-lactamase (CTX-M-3 like) from India and gene association with insertion sequence IS*Ecp1*. FEMS Microbiology Letters. 2001 Jul 1;201(2):237-41.

63. Lartigue MF, Poirel L, Aubert D, Nordmann P. In vitro analysis of IS*Ecp1B*-mediated mobilization of naturally occurring β-lactamase gene blaCTX-M of *Kluyvera ascorbata.* Antimicrobial Agents and Chemotherapy. 2006 Apr 1;50(4):1282-6.

64. Eckert C, Gautier V, Saladin-Allard M, Hidri N, Verdet C, Ould-Hocine Z, Barnaud G, Delisle F, Rossier A, Lambert T, Philippon A. Dissemination of CTX-M-type β-lactamases among clinical isolates of *Enterobacteriaceae* in Paris, France. Antimicrobial Agents and Chemotherapy. 2004 Apr 1;48(4):1249-55.

65. Abbassi MS, Torres C, Achour W, Vinué L, Sáenz Y, Costa D, Bouchami O, Hassen AB. Genetic characterisation of CTX-M-15-producing *Klebsiella pneumoniae* and *Escherichia coli* strains isolated from stem cell transplant patients in Tunisia. International Journal of Antimicrobial Agents. 2008 Oct 1;32(4):308-14.

66. Cullik A, Pfeifer Y, Prager R, von Baum H, Witte W. A novel IS26 structure surrounds blaCTX-M genes in different plasmids from German clinical *Escherichia coli* isolates. Journal of Medical Microbiology. 2010 May 1;59(5):580-7.

67. Lartigue MF, Poirel L, Nordmann P. Diversity of genetic environment of blaCTX-M genes. FEMS Microbiology Letters. 2004 May 1;234(2):201-7.

68. Wellington EM, Boxall AB, Cross P, Feil EJ, Gaze WH, Hawkey PM, Johnson-Rollings AS, Jones DL, Lee NM, Otten W, Thomas CM. The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. The Lancet Infectious Diseases. 2013 Feb 28;13(2):155-65.

69. Zujić Atalić V, Bedenić B, Kocsis E, Mazzariol A, Sardelić S, Barišić M, Plečko V, Bošnjak Z, Mijač M, Jajić I, Vranić-Ladavac M. Diversity of carbapenemases in clinical isolates of *Enterobacteriaceae* in Croatia—the results of a multicentre study. Clinical Microbiology and Infection. 2014 Nov 1;20(11):O894-903.

70. Drinkovic D, Morris AJ, Dyet K, Bakker S, Heffernan H. Plasmid-mediated AmpC beta-lactamase-producing *Escherichia coli* causing urinary tract infection in the Auckland community likely to be resistant to commonly prescribed antimicrobials. The New Zealand Medical Journal. 2015 Mar 13;128(1410):50-9.

71. Vinué L, Saenz Y, Martinez S, Somalo S, Moreno MA, Torres C, Zarazaga M. Prevalence and diversity of extended-spectrum β-lactamases in faecal *Escherichia coli* isolates from healthy humans in Spain. Clinical Microbiology and Infection. 2009 Oct 1;15(10):954-6.

72. Barguigua A, El Otmani F, Talmi M, Zerouali K, Timinouni M. Prevalence and types of extended spectrum β-lactamases among urinary *Escherichia coli* isolates in Moroccan community. Microbial Pathogenesis. 2013 Sep 30;61:16-22.

73. Park SH, Byun JH, Choi SM, Lee DG, Kim SH, Kwon JC, Park C, Choi JH, Yoo JH. Molecular epidemiology of extended-spectrum β-lactamase-producing *Escherichia coli* in the community and hospital in Korea: emergence of ST131 producing CTX-M-15. BMC Infectious Diseases. 2012 Jun 29;12(1):1.

74. Struelens MJ, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Eurosurveillance.2013 Jan 24;18(4):e20386.

75. Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends in Microbiology. 2014 May 1;22(5):282-91.

76. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016 Jan 1;107(1):1-8.

77. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology. 2015 Apr 1;47(3):199-210.

78. Hapuarachchi I. Evaluation of ESBL *E. coli* urinary isolates in Dunedin. University of Otago; 2016.

79. An introduction to Next-Generation Sequencing Technology [Internet]. Illumina,Inc.; 2017 [cited 2019 Jun 3]. Available from: www.illumina.com/technology/next-generation-sequencing.html

80. Illumina Sequencing Technology Highest data accuracy, simple workflow, and a broad range of applications. 2010 Illumina, Inc; 2010.

81. Kluytmans-van den Bergh MF, Rossen JW, Bruijning-Verhagen PC, Bonten MJ, Friedrich AW, Vandenbroucke-Grauls CM, Willems RJ, Kluytmans JA. Whole genome multilocus sequence typing of extended-spectrum beta-lactamase-producing *Enterobacteriaceae*. Journal of Clinical Microbiology. 2016 Sep 14:JCM-01648.

82. Magi A, Semeraro R, Mingrino A, Giusti B, D'aurizio R. Nanopore sequencing data analysis: state of the art, applications and challenges. Briefings in Bioinformatics. 2017 Jun 16;19(6):1256-72.

83. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics Proteomics Bioinformatics [Internet]. 2016 Oct;14(5):265–79. Available from: http://www.sciencedirect.com/science/article/pii/S1672022916301309.

84. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics. 2015 Oct 1;13(5):278-89.

85. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. Genomics, Proteomics & Bioinformatics. 2015 Feb 1;13(1):4-16.

86. Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, Gundlach JH. Nanopore DNA sequencing with MspA. Proceedings of the National Academy of Sciences. 2010 Sep 14;107(37):16060-5.

87. Venkatesan BM, Bashir R. Nanopore sensors for nucleic acid analysis. Nature Nanotechnology. 2011 Oct;6(10):615.

88. 2008 - 2018 Oxford Nanopore Technologies. Types of nanopores [Internet]. Oxford Nanopore Technologies. [cited 2019 Jul 3]. Available from: https://nanoporetech.com/how-it-works/types-of-nanopores#search&

89. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. Nature Biotechnology. 2016 May;34(5):518.

90. Brown CG, Clarke J: Nanopore development at Oxford Nanopore. Nature Biotechnology. 2016;34(8):810–811.

91. Goyal P, Krasteva PV, Van Gerven N, Gubellini F, Van den Broeck I, Troupiotis-Tsaïlaki A, Jonckheere W, Péhau-Arnaudet G, Pinkner JS, Chapman MR, Hultgren SJ. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature. 2014 Dec;516(7530):250.

92. Jefferson OA, Jaffe A, Ashton D, Warren B, Koellhofer D, Dulleck U, Ballagh A, Moe J, DiCuccio M, Ward K, Bilder G. The long view on sequencing. Nature Biotechnology. 2018 Aug 1;36(8):287.

93. Carter JM, Hussain S. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. Wellcome Open Research. 2017 Apr 6;2:23.

94. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, Corbett CR. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. Scientific Reports. 2018 Jul 19;8(1):10931.

95. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nature Protocols. 2017 Jun;12(6):1261.

96. Yamagishi J, Runtuwene LR, Hayashida K, Mongan AE, Thi LA, Thuy LN, Nhat CN, Limkittikul K, Sirivichayakul C, Sathirapongsasuti N, Frith M. Serotyping dengue virus with isothermal amplification and a portable sequencer. Scientific Reports. 2017 Jun 14;7(1):3510.

97. Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, Martellaro C, Falzarano D, Marzi A, Squires RB, Wollenberg KR. Nanopore sequencing as a rapidly deployable Ebola outbreak tool. Emerging Infectious Diseases. 2016 Feb;22(2):331.

98. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research. 2017 May 1;27(5):722-36.

99. Judge K, Hunt M, Reuter S, Tracey A, Quail MA, Parkhill J, Peacock SJ. Comparison of bacterial genome assembly software for MinION data and their applicability to medical microbiology. Microbial Genomics. 2016 Sep;2(9).

100. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, Mathers A, Norris R, Navickaite I, Eaton C, Iqbal Z. Resolving plasmid structures in *Enterobacteriaceae* using the MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. Microbial Genomics. 2017 Aug;3(8).

101. Goldstein S, Beka L, Graf J, Klassen JL. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. BMC Genomics. 2019 Dec;20(1):23.

102. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biology. 2018 Dec;19(1):90.

103. David M, Dursi LJ, Yao D, Boutros PC, Simpson JT. Nanocall: an open source basecaller for Oxford Nanopore sequencing data. Bioinformatics. 2016 Sep 10;33(1):49-55.

104. Boža V, Brejová B, Vinař T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. PloS One. 2017 Jun 5;12(6).

105. Teng H, Cao MD, Hall MB, Duarte T, Wang S, Coin LJ. Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. GigaScience. 2018 Apr 10;7(5).

106. Wick RR, Judd LM, Holt KE. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLoS Computational Biology. 2018 Nov 20;14(11).

107. Wick R. Porechop [Internet]. Available from: https://github.com/rrwick/Porechop

108. Simpson J. Nanopolish [Internet]. Available from: https://github.com/jts/nanopolish

109. de Lannoy C, de Ridder D, Risse J. The long reads ahead: de novo genome assembly using the MinION. F1000Research. 2017 Dec 12;6:1083.

110. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology. 2012 May 1;19(5):455-77.

111. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95.

112. Sovic I, Vaser R, Sikic M. Racon [Internet]. Available from: https://github.com/isovic/racon

113. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS One. 2014 Nov 19;9(11).

114. Boyle JF, Soumakis SA, Rendo A, Herrington JA, Gianarkis DG, Thurberg BE, Painter BG. Epidemiologic analysis and genotypic characterization of a nosocomial outbreak of vancomycin-resistant *Enterococci*. Journal of Clinical Microbiology. 1993 May 1;31(5):1280-5.

115. Moran RA, Anantham S, Holt KE, Hall RM. Prediction of antibiotic resistance from antibiotic resistance genes detected in antibiotic-resistant commensal *Escherichia coli* using PCR or WGS. Journal of Antimicrobial Chemotherapy. 2017 Mar 1;72(3):700-4.

116. European Committee on Antimicrobial Susceptibility Testing. EUCAST guidelines for detection of resistance mechanisms and specific resistances of clinical and/or epidemiological importance. European Society of Clinical Microbiology and Infectious Diseases (EUCAST), Basel, Switzerland. 2016. Available from: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.pdf.

117. Dallenne C, Da Costa A, Decré D, Favier C, Arlet G. Development of a set of multiplex PCR assays for the detection of genes encoding important β-lactamases in *Enterobacteriaceae*. Journal of Antimicrobial Chemotherapy. 2010 Jan 12;65(3):490-5.

118. Lampkowska J, Feld L, Monaghan A, Toomey N, Schjørring S, Jacobsen B, van der Voet H, Andersen SR, Bolton D, Aarts H, Krogfelt KA. A standardized conjugation protocol to asses antibiotic resistance transfer between lactococcal species. International journal of Food Microbiology. 2008 Sep 30;127(1-2):172-5.

119. Woodall CA. DNA transfer by bacterial conjugation. *E. coli* Plasmid Vectors: Methods and Applications. Methods in Molecular Biology. 2003:61-5.

120. Kuhnert P, Nicolet J, Frey J. Rapid and accurate identification of *Escherichia coli* K-12 strains. Applied and Environmental Microbiology. 1995 Nov 1;61(11):4135-9.

121. Wiegand I, Hilpert K, Hancock RE. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature Protocols. 2008 Feb 1;3(2):163.

122. Johnson JR, Johnston B, Kuskowski MA, Colodner R, Raz R. Spontaneous conversion to quinolone and fluoroquinolone resistance among wild-type *Escherichia coli* isolates in relation to phylogenetic background and virulence genotype. Antimicrobial Agents and Chemotherapy. 2005 Nov 1;49(11):4739-44.

123. Heringa SD, Monroe JD, Herrick JB A simple, rapid method for extracting large plasmid DNA from bacteria. Nature Precedings. 2007. Available from: http://dx.doi.org/10.1038/npre.2007.1249.1031.

124. Thermo Scientific NanoDrop™ 1000 Spectrophotometer. T009-TECHNICAL BULLETIN. Thermo Scientific. Available from: http://www.nanodrop.com/Library/T009-NanoDrop%201000-&-NanoDrop%208000-Nucleic-Acid-Purity-Ratios.pdf.

125. Thermo Scientific NanoDrop Spectrophotometers. T042-TECHNICAL BULLETIN. Thermo Scientific. Available from: http://www.nanodrop.com/Library/T042-NanoDrop-Spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf.

126. Thoma S, Schobert M. An improved *Escherichia coli* donor strain for diparental mating. FEMS Microbiology Letters. 2009 Apr 9;294(2):127-32.

127. Sheng Y, Mancino V, Birren B. Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Research. 1995 Jun 11;23(11):1990-6.

128. Bio-Rad Laboratories, Inc. Gene Pulser Xcell™ Electroporation System: Instruction Manual. Life Science Group.

129. Cell Signalling Technology. 6-Tube Magnetic Separation Rack [Internet]. [cited 2019 Mar 19]. Available from: https://media.cellsignal.com/products/images/6185358/6528337/magnetic_rack.jpg

130. Oxford Nanopore Technologies. Getting Started guide: Rapid Sequencing Kit [Internet]. Available from: https://community.nanoporetech.com/guides/minion/rapid/introduction

131. Seemann T, Goncalves da Silva A, Bulach DM, Schultz MB, Kwong JC, Howden BP. Nullarbor Github https://github.com/tseemann/nullarbor.

132. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20.

133. Seemann T (2016) https://github.com/tseemann/abricate.

134. Seemann T (2016) https://github.com/tseemann/snippy.

135. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one. 2010 Mar 10;5(3).

136. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 Jul 20;31(22):3691-3.

137. Alder G, Benson D. Own diagram made with:draw.io [Internet]. Available from: draw.io

138. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016 Mar 19;32(14):2103-10.

139. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Computational Biology. 2017 Jun 8;13(6).

140. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012 Apr;9(4):357.

141. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015 Jun 22;31(20):3350-2.

142 Carattoli A, Zankari E, Garcìa-Fernandez A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. PlasmidFinder and pMLST: in silico detection and typing of plasmids. Antimicrobial Agents and Chemotherapy. 2014 Apr 28.

143. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Research. 2006 Jan 1;34(suppl_1):D32-6. Database available from: http://www-is.biotoul.fr.

144. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Research. 2016 Oct 25.

145. McArthur AG, Wright GD. Bioinformatics of antimicrobial resistance in the age of molecular epidemiology. Current Opinion in Microbiology. 2015 Oct 1;27:45-50.

146. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L. The comprehensive antibiotic resistance database. Antimicrobial Agents and Chemotherapy. 2013 Jul 1;57(7):3348-57.

147. Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21097/.

148. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy. 2012 Jul 10;67(11):2640-4.

149. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research. 2016 Apr 19;44(W1):W242-5.

150. Jolley KA, Bray JE, Maiden MC. Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. Wellcome open research. 2018;3.

151. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, Harbarth S, Hindler JF, Kahlmeter G, Olsson-Liljequist B, Paterson DL. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clinical Microbiology and Infection. 2012 Mar;18(3):268-81.

152. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology. 2014 Mar;15(3):R46.

153. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification and assembly. Briefings in Bioinformatics. 2017 Sep 23.

154. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, Sokurenko E, Price LB, Johnson JR. Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. MSphere. 2016 Aug 31;1(4).

155.Kawamura K, Hayashi K, Matsuo N, Kitaoka K, Kimura K, Wachino JI, Kondo T, Iinuma Y, Murakami N, Fujimoto S, Arakawa Y. Prevalence of CTX-M-Type Extended-Spectrum β-Lactamase-Producing *Escherichia coli* B2-O25-ST131 H 30R Among Residents in Nonacute Care Facilities in Japan. Microbial Drug Resistance. 2018 May 23.

156.Peirano G, van der Bij AK, Gregson DB, Pitout JD. Molecular epidemiology over an eleven-year period (2000-10) of Extended-spectrum β-lactamase-producing *Escherichia coli* causing bacteraemia in a centralized Canadian region. Journal of Clinical Microbiology. 2011 Dec 7:JCM-06025.

157.Yasir M, Ajlan AM, Shakil S, Jiman-Fatani AA, Almasaudi SB, Farman M, Baazeem ZM, Baabdullah R, Alawi M, Al-Abdullah N, Ismaeel NA. 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. 2018 Oct 4;8(1):14847.

158. Alghoribi MF, Gibreel TM, Farnham G, Al Johani SM, Balkhy HH, Upton M. Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic *Escherichia coli* clones in Riyadh, Saudi Arabia. Journal of Antimicrobial Chemotherapy. 2015 Jul 16;70(10):2757-62.

159. Roer L, Hansen F, Thomsen MC, Knudsen JD, Hansen DS, Wang M, Samulioniené J, Justesen US, Røder BL, Schumacher H, Østergaard C. WGS-based surveillance of third-generation cephalosporin-resistant *Escherichia coli* from bloodstream infections in Denmark. Journal of Antimicrobial Chemotherapy. 2017 Mar 22;72(7):1922-9.

160. Schaufler K, Nowak K, Düx A, Semmler T, Villa L, Kourouma L, Bangoura K, Wieler LH, Leendertz FH, Guenther S. Clinically Relevant ESBL-Producing *K. pneumoniae* ST307 and *E. coli* ST38 in an Urban West African Rat Population. Frontiers in Microbiology. 2018 Feb 9;9:150.

161. Müller A, Stephan R, Nüesch-Inderbinen M. Distribution of virulence factors in ESBL-producing *Escherichia coli* isolated from the environment, livestock, food and humans. Science of the Total Environment. 2016 Jan 15;541:667-72.

162. Zurfluh K, Nüesch-Inderbinen M, Morach M, Berner AZ, Hächler H, Stephan R. Extended-spectrum ß-lactamase-producing-*Enterobacteriaceae* in vegetables imported from the Dominican Republic, India, Thailand and Vietnam. Applied and Environmental Microbiology. 2015 Feb 27:AEM-00258.

163. Bubpamala J, Khuntayaporn P, Thirapanmethee K, Montakantikul P, Santanirand P, Chomnawang MT. Phenotypic and genotypic characterizations of extended-spectrum beta-lactamase-producing *Escherichia coli* in Thailand. Infection And Drug Resistance. 2018;11:2151.

164. Chowdhury PR, McKinnon J, Liu M, Djordjevic SP. Multidrug resistant uropathogenic *Escherichia coli* ST405 with a novel, composite IS26 transposon in a unique chromosomal location. Frontiers in Microbiology. 2018;9.

165. Matsumura Y, Yamamoto M, Nagao M, Ito Y, Takakura S, Ichiyama S. Association of fluoroquinolone resistance, virulence genes, and IncF plasmids with extended-spectrum-β-lactamase-producing *Escherichia coli* sequence type 131 (ST131) and ST405 clonal groups. Antimicrobial Agents and Chemotherapy. 2013 Oct 1;57(10):4736-42.

166. Bai L, Wang L, Yang X, Wang J, Gan X, Wang W, Xu J, Chen Q, Lan R, Fanning S, Li F. Prevalence and molecular characteristics of extended-spectrum β-lactamase genes in *Escherichia coli* isolated from diarrheic patients in China. Frontiers in Microbiology. 2017 Feb 13;8:144.

167. Sasaki Y, Taketomo N, Sasaki T. Factors affecting transfer frequency of pAM beta 1 from *Streptococcus faecalis* to *Lactobacillus plantarum*. Journal of Bacteriology. 1988 Dec 1;170(12):5939-42.

168. Wu N, Matand K, Kebede B, Acquaah G, Williams S. Enhancing DNA electrotransformation efficiency in *Escherichia coli* DH10B electrocompetent cells. Electronic Journal of Biotechnology. 2010 Sep;13(5):21-2.

169. Nováková J, Izsáková A, Grivalský T, Ottmann C, Farkašovský M. Improved method for high-efficiency electrotransformation of *Escherichia coli* with the large BAC plasmids. Folia Microbiologica. 2014 Jan 1;59(1):53-61.

170. Kuznetsova O, Ronan R, Pugh J, Dokos R. MinKnow 2.0 [Internet]. Oxford Nanopore Technologies. 10may2018 [cited 2019 Dec 4]. Available from: https://community.nanoporetech.com/posts/minknow-2-0

171. Kuznetsova O, Ronan R, Pugh J, Dokos R, Kent T. MinKnow 2.1 [Internet]. Oxford Nanopore Technologies. 2018 Jul 04. [cited 2019 Dec 4]. Available from: https://community.nanoporetech.com/posts/minknow-2-1-for-minion.

172. Wick RR, Judd LM, Holt KE. Comparison of Oxford Nanopore basecalling tools. January. https://doi. org. 2018;10.

173. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. bioRxiv. 2019 Jan 1:543439.].

174. Wee Y, Bhyan SB, Liu Y, Lu J, Li X, Zhao M. The bioinformatics tools for the genome assembly and analysis based on third-generation sequencing. Briefings in Functional Genomics. 2018 Nov 21;18(1):1-2.

175 . Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics. 2015 Nov 20;32(7):1009-15.

176. Laczny CC, Galata V, Plum A, Posch AE, Keller A. Assessing the heterogeneity of in silico plasmid predictions based on whole-genome-sequenced clinical isolates. Briefings in Bioinformatics. 2017 Dec 5.

177. De Maio N, Shaw LP, Hubbard A, George S, Sanderson N, Swann J, Wick R, AbuOun M, Stubberfield E, Hoosdally SJ, Crook DW. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. BioRxiv. 2019 Jan 1:530824.

178. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM, Dodd R. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Medicine. 2015 Dec;7(1):99.

179. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Feb 19;29(8):1072-5.

180. Shahid M, Singh A, Sobia F, Rashid M, Malik A, Shukla I, Khan HM. blaCTX-M, blaTEM, and blaSHV in *Enterobacteriaceae* from North-Indian tertiary hospital: high occurrence of combination genes. Asian Pacific Journal of Tropical Medicine. 2011 Feb 1;4(2):101-5.

181. Kiiru J, Kariuki S, Goddeeris BM, Butaye P. Analysis of β-lactamase phenotypes and carriage of selected β-lactamase genes among *Escherichia coli* strains obtained from Kenyan patients during an 18-year period. BMC Microbiology. 2012 Dec;12(1):155.

182. Cicek AC, Saral A, Duzgun AO, Yasar E, Cizmeci Z, Balci PO, Sari F, Firat M, Altintop YA, Sibel AK, Caliskan A. Nationwide study of *Escherichia coli* producing extended-spectrum β-lactamases TEM, SHV and CTX-M in Turkey. The Journal of Antibiotics. 2013 Nov;66(11):647.

183. Evans BA, Amyes SG. OXA β-lactamases. Clinical Microbiology Reviews. 2014 Apr 1;27(2):241-63.

184. Palzkill TI, Botstein DA. Identification of amino acid substitutions that alter the substrate specificity of TEM-1 beta-lactamase. Journal of Bacteriology. 1992 Aug 1;174(16):5237-43.

185. Guenther S, Semmler T, Stubbe A, Stubbe M, Wieler LH, Schaufler K. Chromosomally encoded ESBL genes in *Escherichia coli* of ST38 from Mongolian wild birds. Journal of Antimicrobial Chemotherapy. 2017 Feb 3;72(5):1310-3.

186. Rodríguez I, Thomas K, Van Essen A, Schink AK, Day M, Chattaway M, Wu G, Mevius D, Helmuth R, Guerra B. Chromosomal location of blaCTX-M genes in clinical isolates of *Escherichia coli* from Germany, The Netherlands and the UK. International journal of Antimicrobial Agents. 2014 Jun 1;43(6):553-7.

187. Hirai I, Fukui N, Taguchi M, Yamauchi K, Nakamura T, Okano S, Yamamoto Y. Detection of chromosomal blaCTX-M-15 in *Escherichia coli* O25b-B2-ST131 isolates from the Kinki region of Japan. International Journal of Antimicrobial Agents. 2013 Dec 1;42(6):500-6.

188. Ferreira JC, Penha Filho RA, Andrade LN, Berchieri Jr A, Darini AL. Detection of chromosomal blaCTX-M-2 in diverse *Escherichia coli* isolates from healthy broiler chickens. Clinical Microbiology and Infection. 2014 Oct 1;20(10):O623-6.

189. Hamamoto K, Ueda S, Toyosato T, Yamamoto Y, Hirai I. High prevalence of chromosomal blaCTX-M-14 in *Escherichia coli* isolates possessing blaCTX-M-14. Antimicrobial Agents and Chemotherapy. 2016 Apr 1;60(4):2582-4.

190. Botelho LA, Kraychete GB, e Silva C, Lapa J, Regis DV, Picão RC, Moreira BM, Bonelli RR. Widespread distribution of CTX-M and plasmid-mediated AmpC β-lactamases in *Escherichia coli* from Brazilian chicken meat. Memórias do Instituto Oswaldo Cruz. 2015 Apr;110(2):249-54.

191. Papagiannitsis CC, Študentová V, Jakubů V, Španělová P, Urbášková P, Žemličková H, Hrabák J, Czech Participants of European Antimicrobial Resistance Surveillance Network. High prevalence of ST131 among CTX-M-producing *Escherichia coli* from com munity-acquired infections, in the Czech Republic. Microbial Drug Resistance. 2015 Feb 1;21(1):74-84.

192. Vignoli R, García-Fulgueiras V, Cordeiro NF, Bado I, Seija V, Aguerrebere P, Laguna G, Araújo L, Bazet C, Gutkind G, Chabalgoity J. Extended-spectrum β-lactamases, transferable quinolone resistance, and virulotyping in extra-intestinal *E. coli* in Uruguay. The Journal of Infection in Developing Countries. 2016 Jan 31;10(01):43-52.

193. Merino I, Hernández-García M, Turrientes MC, Pérez-Viso B, López-Fresneña N, Diaz-Agero C, Maechler F, Fankhauser-Rodriguez C, Kola A, Schrenzel J, Harbarth S. Emergence of ESBL-producing *Escherichia coli* ST131-C1-M27 clade colonizing patients in Europe. Journal of Antimicrobial Chemotherapy. 2018 Aug 14;73(11):2973-80.

194. Chattaway MA, Jenkins C, Ciesielczuk H, Day M, DoNascimento V, Day M, Rodríguez I, van Essen-Zandbergen A, Schink AK, Wu G, Threlfall J. Evidence of evolving extraintestinal enteroaggregative *Escherichia coli* ST38 clone. Emerging Infectious Diseases. 2014 Nov;20(11):1935.

195. Hagel S, Makarewicz O, Hartung A, Weiß D, Stein C, Brandt C, Schumacher U, Ehricht R, Patchev V, Pletz MW. ESBL colonization and acquisition in a hospital population: The molecular epidemiology and transmission of resistance genes. PloS One. 2019 Jan 14;14(1).

196. Frank T, Gautier V, Talarmin A, Bercion R, Arlet G. Characterization of sulphonamide resistance genes and class 1 integron gene cassettes in *Enterobacteriaceae*, Central African Republic (CAR). Journal of Antimicrobial Chemotherapy. 2007 Mar 9;59(4):742-5.

197. Musicha P, Feasey NA, Cain AK, Kallonen T, Chaguza C, Peno C, Khonga M, Thompson S, Gray KJ, Mather AE, Heyderman RS. Genomic landscape of extended-spectrum β-lactamase resistance in *Escherichia coli* from an urban African setting. Journal of Antimicrobial Chemotherapy. 2017 Mar 4;72(6):1602-9.

198. Xu Y, Sun H, Bai X, Fu S, Fan R, Xiong Y. Occurrence of multidrug-resistant and ESBL-producing atypical enteropathogenic *Escherichia coli* in China. Gut Pathogens. 2018 Dec;10(1):8.

199. Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM, Djordjevic SP. Genomic microbial epidemiology is needed to comprehend the global problem of antibiotic resistance and to improve pathogen diagnosis. Frontiers in Microbiology. 2016 Jun 15;7:843.

200. Ho PL, Leung LM, Chow KH, Lai EL, Lo WU, Ng TK. Prevalence of aminoglycoside modifying enzyme and 16S ribosomal RNA methylase genes among aminoglycoside-resistant *Escherichia coli* isolates. Journal of Microbiology, Immunology and Infection. 2016 Feb 1;49(1):123-6.

201. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2015. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net).Stockholm: ECDC; 2017.

202. Bouchillon S, Hoban DJ, Badal R, Hawser S. Fluoroquinolone resistance among gram-negative urinary tract pathogens: global smart program results, 2009-2010. The Open Microbiology Journal. 2012;6:74.

203. Ekwanzala MD, Dewar JB, Kamika I, Momba MN. Systematic review in South Africa reveals antibiotic resistance genes shared between clinical and environmental settings. Infection and Drug Resistance. 2018;11:1907.

204. Shams F, Hasani A, Rezaee MA, Nahaie MR, Hasani A, Haghi MH, Pormohammad A, Arbatan AE. Carriage of class 1 and 2 integrons in quinolone, extended-spectrum-β-lactamase-producing and multidrug resistant *E. coli* and *K. pneumoniae*: High burden of antibiotic resistance. Advanced Pharmaceutical Bulletin. 2015 Sep;5(3):335.

205. Mazzariol A, Bazaj A, Cornaglia G. Multi-drug-resistant Gram-negative bacteria causing urinary tract infections: a review. Journal of Chemotherapy. 2017 Dec 22;29(sup1):2-9.

206. Robicsek A, Jacoby GA, Hooper DC. The worldwide emergence of plasmid-mediated quinolone resistance. The Lancet Infectious Diseases. 2006 Oct 1;6(10):629-40.

207. Shaheen BW, Nayak R, Foley SL, Boothe DM. Chromosomal and plasmid-mediated fluoroquinolone resistance mechanisms among broad-spectrum-cephalosporin-resistant *Escherichia coli* isolates recovered from companion animals in the USA. Journal of Antimicrobial Chemotherapy. 2013 Jan 9;68(5):1019-24.

208. Paltansing S, Kraakman ME, Ras JM, Wessels E, Bernards AT. Characterization of fluoroquinolone and cephalosporin resistance mechanisms in *Enterobacteriaceae* isolated in a Dutch teaching hospital reveals the presence of an *Escherichia coli* ST131 clone with a specific mutation in parE. Journal of Antimicrobial Chemotherapy. 2012 Sep 18;68(1):40-5.).

209. Vardakas KZ, Legakis NJ, Triarides N, Falagas ME. Susceptibility of contemporary isolates to fosfomycin: a systematic review of the literature. International Journal of Antimicrobial Agents. 2016 Apr 1;47(4):269-85.

210. Ho PL, Chan J, Lo WU, Law PY, Li Z, Lai EL, Chow KH. Dissemination of plasmid-mediated fosfomycin resistance fosA3 among multidrug-resistant *Escherichia coli* from livestock and other animals. Journal of Applied Microbiology. 2013 Mar;114(3):695-702.

211. Nordmann P, Poirel L. Plasmid-mediated colistin resistance: an additional antibiotic resistance menace. Clinical Microbiology and Infection. 2016 May 1;22(5):398-400.

212. Bajaj P, Singh NS, Virdi JS. *Escherichia coli* β-lactamases: what really matters. Frontiers in Microbiology. 2016 Mar 30;7:417.

213. Yang QE, Sun J, Li L, Deng H, Liu BT, Fang LX, Liu YH. IncF plasmid diversity in multi-drug resistant *Escherichia coli* strains from animals in China. Frontiers in Microbiology. 2015 Sep 22;6:964.

214. Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, Crook D, Walker AS, Woodford N, Anjum MF, Stoesser N. Ordering the mob: Insights into replicon and MOB typing schemes from analysis of a curated dataset of publicly available plasmids. Plasmid. 2017 May 1;91:42-52.

215. Burke L, Humphreys H, Fitzgerald-Hughes D. The molecular epidemiology of resistance in cefotaximase-producing *Escherichia coli* clinical isolates from Dublin, Ireland. Microbial Drug Resistance. 2016 Oct 1;22(7):552-.

216. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V. Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. MBio. 2016 May 4;7(2).

217. Rozwandowicz M, Brouwer MS, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. Plasmids carrying antimicrobial resistance genes in *Enterobacteriaceae*. Journal of Antimicrobial Chemotherapy. 2018 Jan 23;73(5):1121-37.

218. Kanamori H, Parobek CM, Juliano JJ, Johnson JR, Johnston BD, Johnson TJ, Weber DJ, Rutala WA, Anderson DJ. Genomic analysis of multidrug-resistant *Escherichia coli* fro m North Carolina community hospitals: ongoing circulation of CTX-M-producing ST131-H30Rx and ST131-H30R1 strains. Antimicrobial Agents and Chemotherapy. 2017 Aug 1;61(8).

219. Guiral E, Pons MJ, Vubil D, Marí-Almirall M, Sigaúque B, Soto SM, Alonso PL, Ruiz J, Vila J, Mandomando I. Epidemiology and molecular characterization of multidrug-resistant *Escherichia coli* isolates harboring blaCTX-M group 1 extended-spectrum β-lactamases causing bacteremia and urinary tract infection in Manhiça, Mozambique. Infection and Drug Resistance. 2018;11:927.

220. Benzerara Y, Gallah S, Hommeril B, Genel N, Decré D, Rottman M, Arlet G. Emergence of plasmid-mediated fosfomycin-resistance genes among *Escherichia coli* isolates, France. Emerging Infectious Diseases. 2017 Sep;23(9):1564.

221. Mo SS, Kristoffersen AB, Sunde M, Nødtvedt A, Norström M. Risk factors for occurrence of cephalosporin-resistant *Escherichia coli* in Norwegian broiler flocks. Preventive Veterinary Medicine. 2016 Aug 1;130:112-8. *Escherichia coli*

222. Iovleva A, Bonomo RA. The ecology of extended-spectrum β-lactamases (ESBLs) in the developed world. Journal of Travel Medicine. 2017 Apr 1;24(suppl_1):S44-51.

223. Zhang L, Lü X, Zong Z. The emergence of bla CTX-M-15-carrying *Escherichia coli* of ST131 and new sequence types in Western China. Annals of Clinical Microbiology And Antimicrobials. 2013 Dec;12(1):35.

224. Kim JW, Mira P, Chan PP, Lowe TM, Barlow M, Camps M. Functionally redundant forms of extended-spectrum beta-lactamases and aminoglycoside modifying enzymes drive the evolution of two distinct multidrug resistance gene clusters in clinical populations of EXPEC. bioRxiv. 2018 Jan 1:367938.

225. Medaney F, Ellis RJ, Raymond B. Ecological and genetic determinants of plasmid distribution in *Escherichia coli*. Environmental Microbiology. 2016 Nov;18(11):4230-9.

226. Pilla G, Tang CM. Going around in circles: virulence plasmids in enteric pathogens. Nature Reviews Microbiology. 2018 May 31:1.

227. Qu F, Ying Z, Zhang C, Chen Z, Chen S, Cui E, Bao C, Yang H, Wang J, Liu C, Mao Y. Plasmid-encoding extended-spectrum β-lactamase CTX-M-55 in a clinical *Shigella sonnei* strain, China. Future Microbiology. 2014 Oct;9(10):1143-50.

228. Yang X, Liu W, Liu Y, Wang J, Lv L, Chen X, He D, Yang T, Hou J, Tan Y, Xing L. F33: A-: B-, IncHI2/ST3, and IncI1/ST71 plasmids drive the dissemination of fosA3 and blaCTX– M–55/– 14/– 65 in *Escherichia coli* from chickens in China. Frontiers in Microbiology. 2014 Dec 16;5:688.

229. Pérez-Etayo L, Berzosa M, González D, Vitas A. Prevalence of Integrons and Insertion Sequences in 6ESBL-Producing *E. coli* Isolated from Different Sources in Navarra, Spain. International Journal of Environmental Research and Public Health. 2018 Oct;15(10):2308.

230. Zhang D, Zhao Y, Feng J, Hu L, Jiang X, Zhan Z, Yang H, Yang W, Gao B, Wang J, Li J. Replicon-based typing of IncI-complex plasmids, and comparative genomics analysis of IncIγ/K1 plasmids. Frontiers in Microbiology. 2019;10:48.

231. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M, Schembri MA. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. PloS One. 2011 Oct 28;6(10):e26578.

232. Liao W, Jiang J, Xu Y, Yi J, Chen T, Su X, Pan S, Wei X, Li Y. Survey for β-lactamase among bacterial isolates from Guangzhou, China hospitals between 2005–2006. The Journal of Antibiotics. 2010 May;63(5):225.

233. Tian SF, Chu YZ, yi Chen B, Nian H, Shang H. ISEcp1 element in association with blaCTX-M genes of *E. coli* that produce extended-spectrum β-lactamase among the elderly in community settings. Enfermedades Infecciosas y Microbiologia Clinica. 2011 Dec 1;29(10):731-4.

234. Hu X, Gou J, Guo X, Cao Z, Li Y, Jiao H, He X, Ren Y, Tian F. Genetic contexts related to the diffusion of plasmid-mediated CTX-M-55 extended-spectrum beta-lactamase isolated from *Enterobacteriaceae* in China. Annals of Clinical Microbiology and Antimicrobials. 2018 Dec;17(1):12.

235. Sun J, Li XP, Yang RS, Fang LX, Huo W, Li SM, Jiang P, Liao XP, Liu YH. Complete nucleotide sequence of an IncI2 plasmid coharboring blaCTX-M-55 and mcr-1. Antimicrobial Agents and Chemotherapy. 2016 Aug 1;60(8):5014-7.

236. Uemura S, Yokota SI, Mizuno H, Sakawaki E, Sawamoto K, Maekawa K, Tanno K, Mori K, Asai Y, Fujii N. Acquisition of a transposon encoding extended-spectrum β-lactamase SHV-12 by Pseudomonas aeruginosa isolates during the clinical course of a burn patient. Antimicrobial Agents and Chemotherapy. 2010 Sep 1;54(9):3956-9.

237. Kassis-Chikhani N, Frangeul L, Drieux L, Sengelin C, Jarlier V, Brisse S, Arlet G, Decré D. Complete nucleotide sequence of the first KPC-2-and SHV-12-encoding IncX plasmid, pKpS90, from *Klebsiella pneumoniae*. Antimicrobial Agents and Chemotherapy. 2013 Jan 1;57(1):618-20.

238. Diestra K, Juan C, Curiao T, Moya B, Miro E, Oteo J, Coque TM, Pérez-Vázquez M, Campos J, Canton R, Oliver A. Characterization of plasmids encoding bla ESBL and surrounding genes in Spanish clinical isolates of *Escherichia coli* and *Klebsiella pneumoniae*. Journal of Antimicrobial Chemotherapy. 2008 Nov 6;63(1):60-6.

239. Amos GC, Gozzard E, Carter CE, Mead A, Bowes MJ, Hawkey PM, Zhang L, Singer AC, Gaze WH, Wellington EM. Validated predictive modelling of the environmental resistome. The ISME Journal. 2015 Jun;9(6):1467.

240. Murray AK, Zhang L, Yin X, Zhang T, Buckling A, Snape J, Gaze WH. Novel Insights into Selection for Antibiotic Resistance in Complex Microbial Communities. MBio. 2018 Sep 5;9(4).

241. Goering RV, Köck R, Grundmann H, Werner G, Friedrich AW. From theory to practice: molecular strain typing for the clinical and public health setting. Eurosurveillance. 2013;18(4). Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20383.

242. Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, Andersson DI. Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathogens. 2011 Jul 21;7(7):e1002158.

# 7 APPENDIX A

- Figure 1

| *Albacore* run on raw signal MinION fast5 output files |
| --- |
| "C:\Program Files\OxfordNanopore\ont-Albacore\read_fast5_basecaller.exe" --<br>flowcell FLO-MIN106 --kit SQK-RBK108 --barcoding --recursive --<br>output_format fast5,fastq --input directory_of_fast5_files --<br>save_path directory_for_output --worker_threads 4 |

- Figure 2

| .sh file written in nano |
| --- |
| nano Minion_read_1.sh<br>for file in *.fastq; do<br>porechop -i $file -b outputfolder<br>done |

- Figure 3

| *Miniasm/Minimap* on MinION long reads |
| --- |
| /APPS/minimap/minimap -Sw5 -L100 -m0 -<br>t8 Minion_reads.fastq Minion_reads.fastq \|gzip -1 > Minion_reads.paf.gz<br>/APPS/miniasm/miniasm -f Minion_reads.fastq Minion_reads.paf.gz > Minion_reads.gfa<br>awk '/^S/{print ">"$2"\n"$3}' Minion_reads.gfa \| fold > Minion_reads.fa<br>grep ">" Minion_reads.fa \| wc -l |
| *SPAdes* on paired *Illumina* short reads and the *Miniasm/Minimap* long read only assembly |
| spades.py -plasmid -1 illumina_shortreads_forward.fastq.gz -<br>2 Illumina_shortreads_reverse.fastq.gz --untrusted-contigs Minion_reads.fa -<br>o miniasm_spades_hybrid_assembly |

- Figure 4

| run *Miniasm/Minimap* on long reads to build long read only assembly |
| --- |
| /APPS/minimap/minimap -Sw5 -L100 -m0 -t8 BC01.fastq BC01.fastq \| gzip -1 > BC01.paf.gz<br>/APPS/miniasm/miniasm -f BC01.fastq BC01.paf.gz > BC01.gfa<br>awk '/^S/{print ">"$2"\n"$3}' BC01.gfa \| fold > BC01.fa<br>grep ">" BC01.fa \| wc -l |
| OR use sed to flatten the fastq to fasta |
| sed '/^@/!d;s//>/;N' BC01.fq > BC01.fa |

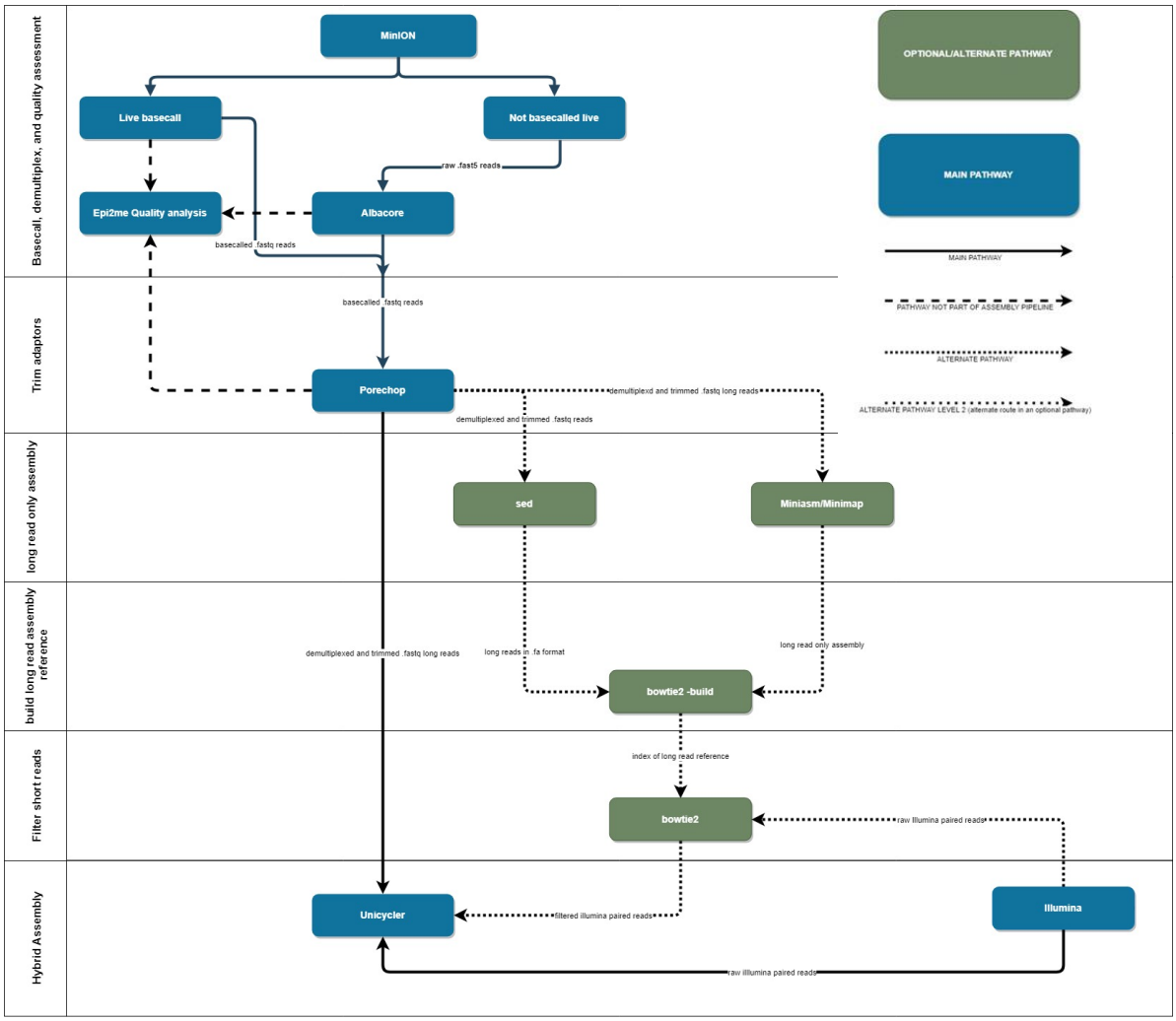
- Figure 5

| *Bowtie2* to build an index |
| --- |
| bowtie2-build -f BC01.fa index_B4R4BC01 |
| Run *bowtie2* to filter reads |
| bowtie2 –q --very-sensitive-local -x index _B4R4BC01 -1 shortread1.fastq.gz -2 shortread2.fastq.gz --al-conc ./mappedshortread.fq -S test.sam --no-unal -N 1 |

- Figure 6

| *Unicycler* command |
| --- |
| Unicycler -mode normal -1 mappedshortread.1.fq -2 mappedshortread.2.fq -l longread.fastq.gz -o outputfolder |

- Figure 7


Basecall, demultiplex, and quality assessment
Trim adaptors
long read only assembly
build long read assembly reference
Filter short reads
Hybrid Assembly
MinION
Live basecall
Not basecalled live
raw .fast5 reads
Epi2me Quality analysis
Albacore
basecalled .fastq reads
basecalled .fastq reads
Porechop
demultiplexd and trimmed .fastq long reads
demultiplexed and trimmed .fastq reads
sed
Miniasm/Minimap
demultiplexed and trimmed .fastq long reads
long reads in .fa format
long read only assembly
bowtie2 -build
index of long read reference
bowtie2
raw illumina paired reads
Unicycler
filtered illumina paired reads
Illumina
raw illumina paired reads
OPTIONAL/ALTERNATE PATHWAY
MAIN PATHWAY
MAIN PATHWAY
PATHWAY NOT PART OF ASSEMBLY PIPELINE
ALTERNATE PATHWAY
ALTERNATE PATHWAY LEVEL 2 (alternate route in an optional pathway)

# 8 APPENDIX B

## SUPPLEMENTARY DATA FILE

Description:

The accompanying HTML file contains the Nullarbor report which shows the sequence assembly metrics and annotation, MLST, species identification, resistome, core genome phylogeny and pan genome phylogeny.

Filename:
Supplemental_data_Nullarbor