Genomic Epidemiology of *Escherichia coli* in Human Blood-Stream Infections

A thesis submitted for the degree of

Doctor of Philosophy

by

Priyanka Shirish Hastak

MSc. (Biotechnology)

Certificate of Original Authorship

I, Priyanka Shirish Hastak, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science, School of Life Sciences at the University of Technology Sydney is wholly my own work unless otherwise referenced or acknowledged.

I certify that all information sources and literature used are indicated in this thesis.

This document has not been submitted to any other academic institution.

This research is supported by an Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Priyanka Shirish Hastak

Date: 3rd March 2021

Acknowledgement:

I would first like to thank my supervisors, Associate Professor Garry Myers, Dr. Piklu Roy Chowdhury and Professor Steven Djordjevic for their guidance, training and support during my candidature. I would like to specially thank them for giving me the opportunity to work on such an excellent research project. It has been an honour to work with them and learn from them in the last 3 and half years. Also, I would like to thank the AusGem (Australian centre for genomic epidemiology microbiology) and Concord Repatriate General Hospital, Sydney for providing me with a great collection of *Escherichia coli* strains that I had the privilege to work with during my candidature.

I would like to thank the University of Technology Sydney and the i3 Institute for supporting me financially with the International Research Scholarship and the i3 Institute Research Scholarship, as well as the Australian Government Research Training Program.

Thank you to all the members of Djordjevic lab, past and present. I would particularly like to thank Ethan Wyrsch, Cameron Reid and Max Cummins for showing me how to work with a number of bioinformatics tools. Also thank you Veronica Jarocki, Dmitriy Li, Tiziana Zingali and Elisa Massella, for being great friends and colleagues that have made my candidature enjoyable and fun.

Thank you to the academic staff and technical staff at the i3 Institute, that includes Professor Liz Harry for her support and encouragement. I am glad and proud to be a part of the i3 Institute that has a friendly environment for young scientists to work on cutting edge research in infectious disease.

Lastly but most importantly, I would like to thank my family and friends. My loving and amazing parents for encouraging and believing in me, always. My sister, Radhika who has supported me spiritually throughout my thesis. Thank you to my aunt and uncle in Sydney for being my home, away from home. Thank you to all my friends back in India and here in Sydney for being my support team during my candidature. Love you all.

Statement:

This thesis is by compilation. The first and second result chapters are publications. The publications are listed in the section below. The last result chapter will be submitted for publishing shortly. The figures, tables and supplementary data is provided at the end of each chapter or as separate files in folder submitted with the thesis.

Publications arising from this thesis

1. Hastak, P.; Cummins, M.L.; Gottlieb, T.; Cheong, E.; Merlino, J.; Myers, G.S.A.; Djordjevic, S.P.; Chowdhury, P.R.Genomic profiling of Escherichia coli isolates from bacteraemia patients: A 3-year cohort study of isolates collected at a Sydney teaching hospital. Microb. Genom. 2020, 6,e000371. https://doi.org/10.1099/mgen.0.000371

CRediT Author statement:

Priyanka Hastak: Methodology, Writing- Original Draft, Sample Collection, Writing- Reviewing and Editing, Data Curation, Visualisation. **Piklu Roy Chowdhury**: Conceptualization, Methodology, Writing- Reviewing and Editing, Validation, Data Curation, Project administration, Supervision, Funding acquisition, Resources. **Garry Myers**: Supervision, Validation, Time Management, Methodology, Project Administration, Writing- Review and Editing. **Steven Djordjevic:** Supervision, Validation, Time Management, Writing- Review and Editing, Funding acquisiton. **Tom Gottleib:** Project administration, Fund acquistion, Resources. **John Merlino**: Methodology, Resources. **Elaine Cheong**: Project administration, Resources. **Max Cummins**: Methodology, Software, Visualisation.

2. Hastak, P., Fourment, M., Darling, A., Gottleib., T, Cheong, E., Merlino, J., Myers, G., Djordjevic, S.P., and Roy Chowdhury, P. *Escherichia coli* ST8196 is a novel, locally evolved, and extensively drug resistant pathogenic lineage within the ST131 clonal complex. Emerg Microbes Infect. 2020 Dec;9(1):1780-1792. doi: 10.1080/22221751.2020.1797541.
CRediT Author statement:

Priyanka Hastak: Methodology, Writing- Original Draft, Writing- Reviewing and Editing, Data Curation, Visualisation. **Piklu Roy Chowdhury**: Conceptualization, Methodology, Writing- Reviewing and Editing, Validation, Data Curation, Project administration, Supervision, Funding acquisition, Resources. **Garry Myers**: Supervision, Validation, Time Management, Methodology, Project Administration, Writing- Review and Editing. **Steven Djordjevic:** Supervision, Validation, Time Management, Writing- Review and Editing, Funding acquisiton. **Aaron Darling**: Conceptualization, Validation, Methodology, Software.

Mathieu Fourment: Methodology, Software, Visualization. **Tom Gottleib:** Project administration, Resources. **John Merlino**: Methodology, Resources. **Elaine Cheong**: Project administration, Resources.

Contents presented in chapter 6 will be included in a manuscript which is in preliminary draft stages.

Hastak, P., Gottlieb, T., Cheong, J., Myers, G., Djordjevic, S.P., and Roy Chowdhury, P.

Comparative genomic analysis of ST38 *Escherichia coli* from human, animal and environmental settings reported in Australia and globally.

Table of Contents

List of abbreviations:

AMR	Antimicrobial Resistance
ARG	Antimicrobial Resistance Gene
MGE	Mobile Genetic Element
IS	Insertion Sequence
DNA	Deoxyribonucleic acid
cc	Clonal complex
bp	Base pair
CRL	Complex Resistance Locus
BSI	Blood-stream infections
ESBL	Extended-spectrum beta-lactamase
ETEC	Enterotoxigenic *Escherichia coli*
ExPEC	Extraintestinal Pathogenic *Escherichia coli*
UTI	Urinary Tract Infection
LCBs	Locally Colinear Blocks
LGT	Lateral Gene Transfer
MDR	Multiple-Drug (Antibiotic) Resistance
MLST	Multi-locus Sequence Typing
ST	Sequence Type
SNP	Single Nucleotide Polymorphism
WGS	Whole Genome Sequence
HO	Hospital Onset
CO	Community Onset
TE	Transposable Elements

GI	Genomic Island
MLST	Multi-Locus Sequence Typing
CS	Conserved Segment
ICU	Intensive Care Unit
VAG	Virulence Associated Genes

Abstract:

Blood-stream infections (BSI) are associated with high mortality and morbidity world-wide. The most common aetiological agent of these infections is *Escherichia coli*. The rise of multi-drug resistant (MDR) *E. coli* is a major concern in clinical medicine and needs to be monitored for implementation of infection control strategies. The misuse of antimicrobials in clinical treatment, and as growth promotors in food-producing animals, is a key component of the rapid evolution of MDR *E. coli*. MDR organisms spread antimicrobial resistance and virulence factors by lateral gene transfer via mobile genetic elements (MGEs). Despite this, there is limited knowledge of the origins and underlying mechanisms of BSI MDR *E. coli* infections in Australia.

We conducted genomic epidemiological analyses of *E. coli* from human blood-cultures collected from a Sydney teaching hospital. The collection was dominated by *E. coli* causing community onset BSI, with clones carrying a plethora of virulence and antimicrobial resistance genes. Clinical class 1 integrons associated with IS*26* were identified in a number of sequence types (ST), indicating that they are important drivers of evolution and spread of clinically important antimicrobial resistance genes. We identified IncFII-IncFIB plasmids in the majority of our collection. A number of STs harboured ColV like IncFII plasmids, carry a number of important virulence traits, were also identified. We observed a novel ST, ST8196, that clustered with globally disseminated *E. coli* ST131 isolates. We also identified an emerging ST, ST38, that exhibited resistance to a broad range of clinical beta-lactamases that have been identified in a number of pandemic *E. coli* isolates.

Our study provides evidence that *E. coli* BSIs in Australia have a reservoir of antimicrobial resistance and virulence determinants that are potentially circulating not only in the community and hospital settings but across agricultural settings. These genes are not only present to clinical settings but can circulate within different ecosystems potentially via plasmids or other MGEs. Our findings also reveal that there are specific antimicrobial resistance genes such as ESBLs that are found in dominant *E. coli* clones in both humans and food producing animals. This is likely to occur due the overuse of antimicrobials in human and animal settings that is driving the increase in antimicrobial resistance among bacteria that cause diseases. It is therefore

essential to rigorously monitor these infections to help manage the global problem of antimicrobial resistance, and to reduce or prevent disease outbreaks.

Chapter 1: Thesis Overview

1.1 Background:

This thesis is a pilot study applying genomic surveillance methods to *Escherichia coli* isolates collected from patients with blood-stream infections (BSI) from an Australian hospital. Isolates were collected from 81 BSI patients.

Metadata pertaining to the source of infection, phenotypic antimicrobial resistance profile was collected. The number of days spent at the hospital prior to isolation of strains was taken into consideration for categorisation as either community or hospital onset infections.

1.2 Aims:

The three major aims of this thesis are:

1. Genomic characterisation of multidrug-resistant *Eschericha coli* isolated from patients treated for BSI over 3 years at the Concord Repatriation Hospital in Sydney, Australia.

2. To identify antimicrobial resistance and virulence carriage in hospital and community onset *E. coli* BSI collected from Concord Repatriate General Hospital from 2013-2016. In addition to identify the variety of sequence types (STs) and determine phylogenetic relatedness. The results and conclusions from this genomic data will provide evidence on a) highly resistant and virulence *E. coli* clonal lineages; b) the genetic diversity seen in *E. coli* BSI and the dominant STs that were prevalent in these infections in Australia.Phylogenomic characterisation of dominant, emerging and novel multidrug-resistant *Eschericha coli* lineages identified in the course of this study.

3. To conduct molecular characterisation of novel and emerging multi-drug resistant STs in *E. coli* BSI to reveal the associated mobile genetic elements (MGE) carrying antimicrobial resistance and virulence genes circulating in these infections. The findings will reveal clinically important information on evolving *E. coli* clones in BSI.

Studies describing whole genome sequence analyses of MDR Gram-negative *E. coli* BSI in Australia (and globally) are limited. This thesis provides insights into antimicrobial resistance

and virulence gene carriage associated with dissemination of MDR genes in *E. coli* BSI. This thesis also defines the population structures and overall distribution of sequence types, phylogroups and serotypes of the cohort.

Thirty different STs are found in this study, including the novel STs, ST8196 and ST8197, as well as globally dominant STs, ST73 and ST131. We identified clustered antibiotic resistance genes associated with clinical class 1 integrons and IS*26* (Chapter 4). We conducted an in depth analysis of ST8196, a novel ST identified in this cohort, and characterised three MDR isolates belonging to a rare ST, ST38 (Chapter 5 and Chapter 6). Furthermore, we conducted a detailed analysis of MDR plasmids with clustered antimicrobial resistance and other virulence factors circulating in this cohort (Chapter 5 and Chapter 6).

Chapter 2: Literature Review

2.1 Antibiotic Resistance: A Global Health Concern

Antibiotic resistance emerged shortly after the introduction of antibiotics in the late 1940s [1]. The subsequent explosive dissemination of antibiotic resistant infections has become a major global threat, resulting in widespread prolonged illness, disability and death. The first antibiotic, penicillin was discovered in 1928 and antibiotics were first prescribed to treat serious infections in the 1940s. However, penicillin resistance rapidly emerged in the 1950s [2]; this was the first sign of the resistance crisis that affects us now. As depicted in **Figure 1A** the introduction of novel antimicrobials continued until the end of the 1970s - the major antimicrobials that were discovered in this time were tetracycline, methicillin, gentamycin, erythromycin [3]. Fluroquinolones and ciprofloxacin were the most widely used antibiotics for gastroenteritis and respiratory infections in the recent years. Their extensive use resulted in rise of fluroquinolone and ciprofloxacin resistance in developing countries and subsequent degradation of their utility in infection treatment

Multi-drug resistant (MDR) bacteria have evolved over time. MDR bacteria are typically resistant to three or more classes of antibiotics. According to the EUCAST (European Committee on Antimicrobial Suseptibility Testing) guidelines, antimicrobial suseptibility profiles can be generated using AES (Advance Expert System) software designed from VITEK 2 (BioMerieux) [4].

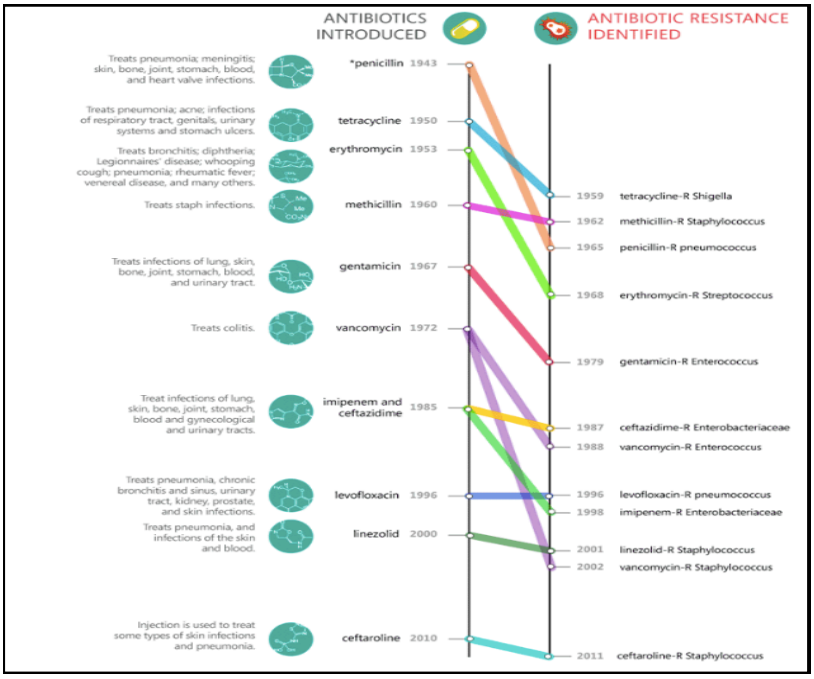


Figure 1A: Timeline for discovery of clinically important antimicrobials and their role in treating specific infections (on the left). Timeline for the emergence of antimicrobial resistance in specific pathogens [2].

Emerging resistance has been particularly identified in tuberculosis and blood-stream infections. Recent global reports indicate that the treatment costs associated with AMR have risen to as high as $2.2 billion dollars annually. Additional treatment costs associated with AMR infections added $1,383 to the cost per patient [5]. A global surveillance study estimated that drug resistant infections will kill an additional 10 million people per year worldwide by 2050 [6]. This study also reported that the most common MDR pathogens were *Klebsiella pneumonia, E. coli* and *Staphylococcus aureus,* of which at least 10% were associated with blood-stream infections [6]. The CDC announced in 2013 that we are now in the "post-antibiotic era" [2], reflecting the urgency of the antibiotic crisis.

Exposure of antibiotics creates selective pressure that can lead to spontaneous bacterial phenotypic changes; genetic mutations in the bacterial chromosome and/or in the plasmids that drives the acquisition of antimicrobial resistance genes via horizontal gene transfer [7]. A US CDC report demonstrated that inappropriate choice and duration of antibiotic treatment in up to 30-50% patients in hospitals resulted in MDR infections that are untreatable or difficult to treat, thus prolonging hospital stays. In addition, 30-60% of antibiotics were inappropriately

administered in intensive care unit patients. This is considered to be a major factor in the persistence of antibiotic resistant strains [8].

Surveillance in Australia indicated that in 21% of patients, antibiotics were misused for surgical prophylaxis and that this correlated with increasing multi-drug resistance within hospital settings [9]. This misuse of antibiotics and inappropriate exposure of patients to antibiotics led to the rapid spread of antimicrobial resistance and thus poorly treated or untreatable infections.

In both developed and under-developed countries, antibiotics are not only widely used in clinical settings but also in livestock or food-producing animals. The antibiotics used in agriculture far exceeds the use in humans [10]. A report in the late 1990s indicate that an average of 700 tonnes of antibiotics is imported into Australia every year, of which 1/3rd is consumed by humans and 550 tonnes used for livestock feed as growth promoters [10] [2]. A significant proportion of all administered antibiotics is excreted in waste. Human sewage and food animal production waste is thus heavily contaminated with residual antibiotics. While much of the waste is processed, the purification systems were not designed to eliminate antimicrobial residues. Hence, large amounts are released into the environment, including prime agricultural lands used to grow fresh produce [11]. In a 2013 study in China, 58,000 tonnes of antimicrobial residues were estimated to be released every year from humans and food animals; moreover the amount of antimicrobials released in SE Asia is staggering [12]. According to a 2015 study in India, a pharmaceutical company manufacturing 90 drugs released 28,000µg/L-30,000µg/L of ciprofloxacin residue every 2 days into nearby water bodies as well as several kilograms of other antimicrobial residues including cetirizine, terbinafine, enoxacin, and norfloxicin [12] [13].

Antibiotic resistance is widely reported in agricultural environments as up to 90% of antibiotics administered to livestock is excreted in the urine and faeces. This is dispersed into soil in the form of fertilizers, waste-water systems and surface run-off [14]. About 80% of the antibiotics in the United States given to livestock are used primarily for growth promotion and for prevention of infection [15]. Most of the antimicrobials used for treating humans are also administered to food-producing animals. This misuse of antibiotics in food-producing animals results in selection of resistant bacterial populations rather than suseptible ones caused by genetic changes, potentially conferring an evolutionary advantage for resistant bacterial

populations. These resistant bacteria can be transferred into humans through food contamination, and can result in severe, untreatable infections.

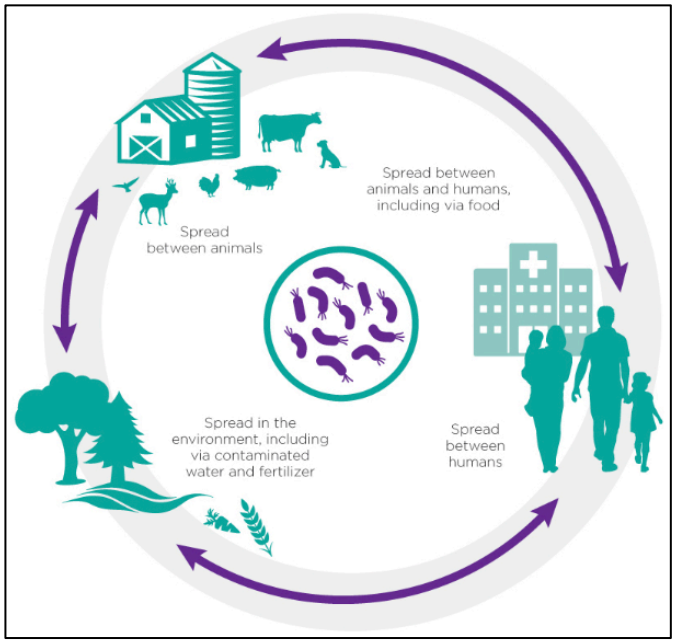


Figure 1B: Route of transmission of antibiotics resistance across food-producing animals, humans and environment [16].

The inappropriate prescribing of antibiotics in human health, combined with the abuse of antibiotics in food-producing animals and their contamination of eocsystems, are the main drivers of antimicrobial resistance world-wide [16] [17]. The spread of antibiotic resistance in human, animal and environment sectors is interdependent. In order to address this issue, health science professionals are now promoting the concept of a "One Health" approach, which proposes a collaborative effort to achieve optimal health for people, animals and environment by taking measures to: (i) prevent infections by reducing over-prescribing antibiotics in human health and promoting effective sanitation methods; (ii) ensure that there is controlled antimicrobial use in food producing animals; and (iii) identify ways to treat industrial and farm waste to mimimise antibiotic ecosystem contamination [16].

2.2 Antimicrobial resistance (AMR) in blood-stream infection (BSI):

The recent emergence of AMR in BSI is a major contributor to high mortality and morbidity in Australia and worldwide. According to Heldens et al., the incidence of these infections is underestimated in Australian intensive care units [20]. BSI pose a significant health burden

with >20-30 million cases occurring every year globally [20]. BSIs are defined by the presence of viable bacteria or fungi in the blood-stream that produces an inflammatory response [18] [19]. These infections are frequently found in critically ill hospitalised ICU patients, and in patients post-surgery or post-diagnostic procedures as these patients often have compromised immune systems [19].

2.2.1 Aetiology of BSI:

BSI disease outcomes stem from a breach of epithelial barriers along the respiratory, gastrointestinal, skin or genital organs by viral or bacterial pathogens that facilitate entry of the foreign organism/s into the bloodstream. BSI spread rapidly throughout the body, as infectious agents can easily access various organs in the body via the bloodstream. This can be fatal as bacterial growth in the blood-stream can progress to sepsis, causing systemic inflammation and abnormal coagulation responses in the host, a condition termed as septicaemia. This results in homeostatic changes that includes impaired immune and endocrine systems and organ damage. Sepsis is a major public health concern leading to 6 million deaths worldwide annually. According to a recent study by the WHO on sepsis, there is incomplete hospital data on the epidemiology of sepsis for many countries, including Australia [20].

The presence of one or more positive blood cultures in a patient indicates a blood-stream infection. These infections are typically divided into three groups. The first group is in patients that are immunologically normal. Some of the pathogens in this category include *Neisseria meningitidis, Streptococcus pyogenes*, and streptococcal BSI in children with native valve endocarditis. The second group includes patients who are physiologically or immunologically impaired, typically the elderly or newborns. The most common bacterial species in this group are *Listeria, E. coli,* and *Klebsiella* spp., and also the fungal pathogen *Candida*. The third group includes patients infected by pathological or pharmacological conditions predisposing to infections. This group includes any Gram-negative or gram-postive pathogens or fungi [19] [21].

The type of pathogens causing BSI has changed over the last two decades, with higher incidence of Gram-negative and fungal infections compared to Gram-positive infections [23]. However, the significant change is not in the bacterial species causing disease *per se*, but rather the antimicrobial resistance patterns, particularly in Gram-negative pathogens. The two main strategies that lead to successful multi-drug or pan-resistant Gram-negative pathogens are (i)

production of extended spectrum beta-lactamases, and (ii) production of carbapenemases that confer resistance to the last resort of antibiotics, that includes imipenem and meropenem [19] [22].

A bacterial strain is defined as multi-drug resistant (MDR) if it is resistant to three or more classes of antibiotics [24]. In Australia, the estimated annual treatment costs associated with MDR BSI reported more than a decade ago were found to be > $686 million dollars [25]. In the United States, the costs to treat Gram-negative BSI ranges from $33,000-44,000 in adult general ICU to $54,000-75,000 in adult surgical ICU [26]. Moreover, a study in Europe reported that treatment costs doubled for patients with MDR Gram-negative BSI [27]. These findings demonstrate that treatment costs for MDR BSI are high, leading to financial burden not only in Australia but globally.

2.2.2 Hospital onset and community onset infections:

Isolates collected from patients after 48 hours from date of admission are deemed as isolates derived from hospital onset (HO) BSI, whereas those collected prior to 48 hours from date of admission are deemed as isolated derived from community onset (CO) BSI. [29].

Hospital onset infections, also known as nosocomial infections, are typically acquired during hospital stays due to contact with other patients, hospital fomites or contaminated hospital equipment. Community onset infections are generally acquired out of the hospital, but may include events like visits to nursing homes and day care facilities and contacts with patients who have been recently discharged from ICUs or health care facilities [28].

However, patients with BSIs moving from ICU facilities to day-care facilities promote the spread of infections across hospital and community settings. Thus the distinction between HO and CO infections is becoming blurred over time [28].

Immunocompromised patients and infected patients with long hospital stays frequently acquire HO infections; cross infections may also occur in intensive care units due to high patient density and prolonged use of urinary catheters. According to a recent surveillance study in Europe, about 7% of patients staying in ICU for more than 2 days acquire hospital infections, of which 4% presented as BSI and 3% UTI. They also revealed that more than 48% of BSI and

98% of UTI are catheter-related infections [30]. These high numbers of HO infections are contributed mainly by antimicrobial resistant pathogens.

As patients within hospitals are typically treated with a variety of antibiotics, the bacterial populations residing within these patients have often been exposed to multiple antibiotics for long periods of time. Under this selective pressure, bacteria can acquire antibiotic resistance genes from the circulating bacterial gene pool in the hospital. This can lead to the dominance of multi-resistant bacterial clones within hospital settings [31].

A recent study identified BSIs as the most frequent hospital-associated infection in patients with urinary tract infections, liver disease and gastro-intestinal diseases. These BSIs were found in more than half of critically ill patients, suggesting that it may be a marker of illness, severity and pre-morbid conditions in patients. A report on ICU patients found that HO-BSI is linked to a 3X mortality risk from the period of positive culture identification [32]. This meta-analysis also revealed that BSI had a directly proportional impact on mortality as higher chances of mortality were observed for 1 in 20 ICU acquired BSI patients. These patients had increased illness severity, surgery, liver or renal disease or mechanical ventilation.

Although HO-BSIs are frequent, there has been a recent escalation of CO-BSI in Australia and worldwide [33]. Antimicrobial resistance within the community is on the rise due to over-crowding and immunocompromised patients visiting community facilities, transmitting these infections to healthy individuals in the community. A surveillance study in Switzerland indicated a sudden increase in community associated BSIs by 14% from 2008 to 2014 [34]. Notably *E. coli* was the most common cause of these infections in Switzerland; a similar increase in *E. coli* CO-BSI was observed in England from 2004-2008 [34] [35].

Primary causes of CO-BSIs include intra-abdominal, gastro-intestinal and respiratory infections. Higher mortality in CO-BSI was associated with inappropriate use of antimicrobials. However, if patients experienced septic shock, then the appropriateness of antimicrobials had less impact on disease progression [36]. The rise in both HO and CO infections may be due to transfer of patients and staff between hospitals and day care facilities. Increased mobility such as patient migration across different geographic locations and international travel may also be a contributing factor [33].

2.2.3 *Escherichia coli* - the most prevalent Gram-negative pathogen in BSI:

In the past two decades, there has been an upsurge of Gram-negative pathogens causing blood-stream infections. MDR Gram-negative BSIs are associated with high mortality rates ranging from 4% to 41% worldwide [37]. These rates differ depending on a variety of risk factors, including age, gender and source of infection, use of catheters, and severity [38]. There are a number of Gram-negative pathogens that cause BSIs in Australia (**Table 1**).

Gram negative pathogen	Incidence of BSI in SA hospitals from 2013-2017
Escherichia coli	41.12%
Pseudomonas aeruginosa	16.80%
Klebsiella spp.	16.06%
Enterobacter spp.	10.75%
Proteus spp.	3.69%
Other Gram-negative bacilli	11.55%

Table 1: List of Gram-negative pathogens causing BSI according to a South Australian (SA) report on BSI [39]

Extra-intestinal pathogenic *E. coli* (ExPEC) have the ability to invade, colonize and cause disease in bodily sites outside the gastrointestinal tract. They are one of the most common causes of BSI in Australia and worldwide [40]. Recent national reports also suggest a rapid rise in fluroquinolone resistance from 13.7% in 2013 to 20.2% in 2016 [41]. As fluroquinolones and cephalosporins are frequently used to treat BSI, frequent administration of these drug classes to combat multi-drug resistant infections has led to an increase in the resistance against these antimicrobials. Consequently the treatment options for BSI are becoming increasingly limited [41].

In the UK, ExPEC BSIs increased by 44% between 2003 and 2011 [43]; more than half of these BSIs were classified as MDR. In Australia, prevalence of MDR *E. coli* causing BSI is as high as 61.6%, which is higher than any other *Enterobacteriaceae* . This report also indicated that the 30 day mortality rate was found in 10.1% cases of *E. coli* bacteremia [42].

MDR *E. coli* BSIs are frequently reported in elderly patients irrespective of gender. Elderly patients that have undergone surgery or are immunocompromised due to cancer or HIV, are

most vulnerable to BSI. Although MDR *E. coli* BSIs are predominately found within hospitalised patients, these infections are estimated to be as high as 55% in community settings in developing countries [44].

A recent study in the United States indicated that there was a nearly equal prevalence of community (48%) and nosocomial (52%) *E. coli* infections [28]. However, in Australia, reports indicate that the proportion of *E. coli* CO-BSI (59%) is higher than HO-BSI (41%) [45]. These numbers suggest a potential pool of AMR genes harboured in these virulent pathogens may be disseminating across and between hospitals and communities, further exacerbating the spread of AMR in these community settings rather than only within hospital settings. The rise in CO-BSI caused by persistant MDR *E. coli* along with travel of patients carrying these infections, could potentially lead to pandemic and/or epidemic outbreaks worldwide.

2.3 Genetic mechanisms that drive the rapid evolution of MDR pathogens:

A major contributor to the rapid dissemination of antimicrobial resistance is lateral gene transfer (LGT). LGT impacts microbial communities and the evolution of successful clones in humans, animals and other environments. It is the movement of DNA between two cells followed by the integration of DNA into the bacterial genome. There are three key mechanisms for promiscuous gene transfer between bacterial cells: transformation, transduction and conjugation [46].

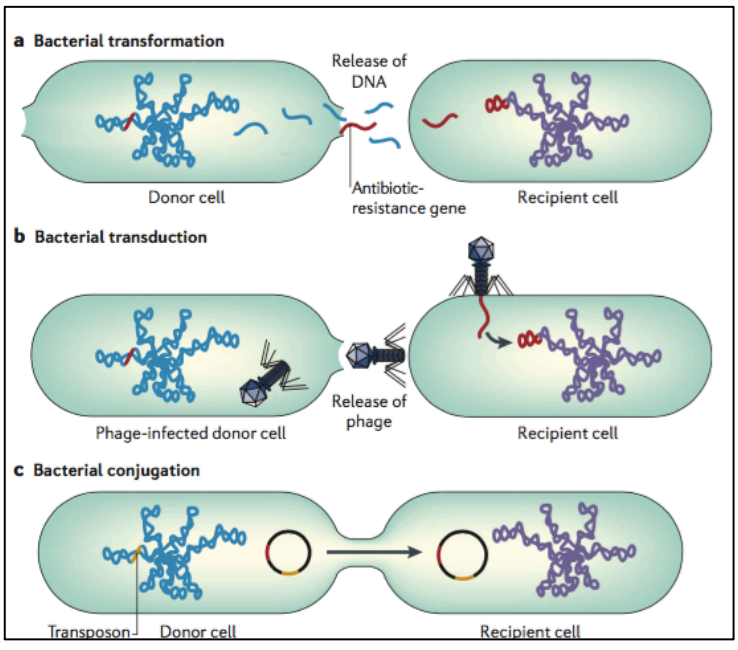


Figure 2: Bacterial gene transfer mechanisms [46]

2.3.1 Transformation:

Transformation enables bacteria to take up, integrate and express extracellular DNA from the environment into a naturally competent recipient cell. Only 60 species of bacteria exhibit natural competency, this includes *Escherichia coli, Klebsiella pneumoniae, Neiseeria spp.,* and *Helicobacter pylori*. The captured DNA may integrate into the competent recipient cell via site specific recombination or homologous recombination, or remain as independently replicating extra-chromosomal entities like plasmids. Homologous recombination occurs more frequently in closely related species compared to divergent bacterial species. Differences in sequences between different species decreases the rate of homologous recombination. In *Escherichia coli*, sexual isolation of species results in sequence divergence.

The incorporation of DNA into the recipient genome of closely related *E. coli* strain can provide an evolutionary advantage, for example, in the evolution and persistance of a specific antibiotic resistant bacterial lineage, as this newly acquired DNA can replicate and transmit into future dominant progeny, helping that lineage to survive in changing environments. [47].

2.3.2 Transduction:

Transduction describes the transfer of DNA from donor bacterium to recipient via phages as vectors. It occurs when newly formed phages take up host genes and transfer them to other bacteria. There are two types of transducing phages; generalized and specialized. Generalised transducting phages can harbour any part of the chromosomal DNA in contrast to specialized transducing phages that harbour only specific fragments of chromosomal DNA. *E. coli* phage P1 was one of the first tranducing phage that was discovered. They observed that bacterial cells get lysed resulting in fragments of chromosomal DNA and the forming phage incorporates the chromosomal DNA into the phage head instead of the phage DNA. The P1 phage is one of the typical examples of generalized transduction as they can transfer any genes of the bacterial donor chromosome [48].

An example of specialized transduction is the λ phage where only specific genes can be transduced. There is an attachment site in λ phage and a recombination site in the bacterial chromosome between the *gal* and *bio* genes for the phage to integrate in the *E. coli* chromosome. Consequent deletion of the bacterial fragments adjacent to the prophage site during transduction may also result in deletion of phage genes. These defected genes are packed

into the phage head and can infect other bacteria. Specialised transduction is restricted to genes in close promixity to integrated prophage, in contrast to generalized transduction that transfers any host gene [48].

Notably, some bacteriophages aquire virulence determinants within their genomes, and lysogenization can convert these strains to a pathogenic variant. For instance, the exotoxin gene A, shiga toxin that are definitive virulence factors responsible for outbreaks and sporadic cases of diarrhea in *E. coli*. This toxin is typically present on bacteriophages that facilitate the transfer of these genes to other bacteria that result in pathogenic strains such as Shiga-toxin producing *E. coli* (STEC) [48] [49].

2.3.3 Conjugation:

Conjugation is plasmid mediated horizontal gene transfer by physical contact between a donor and the recipient bacterium via a mating bridge called the pilus. [50]. In *Escherichia coli*, circular DNA plasmids called F plasmids carry fertility genes that are essential for horizontal gene transfer found in some but not all *E. coli*. In the early 1950s most studies are based on F plasmids and its role in gene transfer. However, in the last few decades the R plasmids are well recognised and these plasmids are typically categorised by their ability to transfer several drug resistance genes by conjugation. R plasmids are known to transfer at lower frequency compared to F plasmids. Notably, some R plasmids when present in the same host were found to inhibit the transfer of F plasmids.

Cells carrying the F plasmid are designated F^+ and those without the F plasmid are designated F^-. The F plasmid can replicate its DNA and allowing the plasmid to be maintained in a dividing cell population. Cells carrying the F plasmid promote synthesis of pili on the bacterial cell surface. The F^+ and F^- cells conjugate, where the F^+ behaves like an F donor and replicates the F plasmid DNA that is then transferred to the F^- recipient. A copy of F remains in the donor cells and the recipient cells gets converted into F^+ cell harbouring a circular F genome. [50].

F plasmids can carry within its genome one or more insertion elements or IS elements. These IS elements are mobile DNA fragments that can move within the host chromosome or between the chromosome and plasmid. The presence of specific IS elements on plasmids and chromosomes results in homologous cross-over at specific sites. When this occurs in two circular DNA the plasmid integrates into the bacterial chromosome. Interestingly when this

occurs F has the ability to transfer the entire host chromosome and the F DNA into the recipient F^- cell. However the process of F factor to integrate into the chromosome occurs only rarely in any cell population. These cells can be isolates and cultivated into pure strains that has the F factor in the chromosome. In these strains each of the cells donate chromosomal allels during F transfer, hence the frequency of recombination is much higher in these cells compared to the normal population [67].

Strains that have integrated F factor into their chromosome are called as high frequency recombination (Hfr) strains. In contrast to these strains the normal F^+ strains harbour only a few rare Hfr cells and therefore depict low frequency of recombination for that strain. The integrated F factor can leave the chromosome and move into the cytoplasm, sometimes also transferring chromosomal genes with it. This F is called as F prime or **F'** and can transfer the host genes into other F^- recipients similar to the manner in which F is spread. Hence the recipient cell will contain two copies of the same gene, one that is present in the bacterial chromosome and the other that has been transferred via the cytoplasmid F prime factor. Hence, these plasmids have the capability to self-transmit between cells and also move fragments of the bacterial chromosome through Hfr recombination.

Plasmids are known to acquire and consequently disseminate mobile genetic eleements such as transposons and IS elements carrying antimicrobial resistance (AMR) genes. The molecular characterisation of plasmids and strain genotypes can provide crucial information on the spread of AMR that can be driven by either pandemic plasmids found in variety of different bacteria from different origins and found in different countries or by spread of bacterial clones that carry these plasmids with AMR genes [67].

2.4 Mobile genetic elements and their role in AMR:

Mobile genetic elements (MGE) are segments of DNA that encode enzymes and other proteins that facilitate the movement of fragments of DNA between different replicating units within the bacterial genome or between bacterial cells [51]. MGE include transposons, plasmids and integrons. One of the clinically important MGEs are integrons. Integrons are site-specific recombination systems that capture arrays of promoter-less genes in specific units called ‘gene cassettes’. These cassettes typically comprise of genes that favour the survival of bacteria and its progeny. They play a crucial role in bacterial evolution and adaption [52]. Although

integrons themselves are immobile, they are located on transferrable plasmids which can spread between and across bacterial species. Integrons acquire mobile antimicrobial resistance gene (ARG) cassettes by site specific recombination. Over 130 different gene cassettes have been previously identified. In clinical settings, integrons are typically associated with 2-6 gene cassettes that encode for clinically important antimicrobial resistance genes. The ability of acquiring these gene cassettes between bacterial strains has resulted in rapid evolving multi-drug resistant bacteria that circulate in humans [53] **(Figure 3)**.

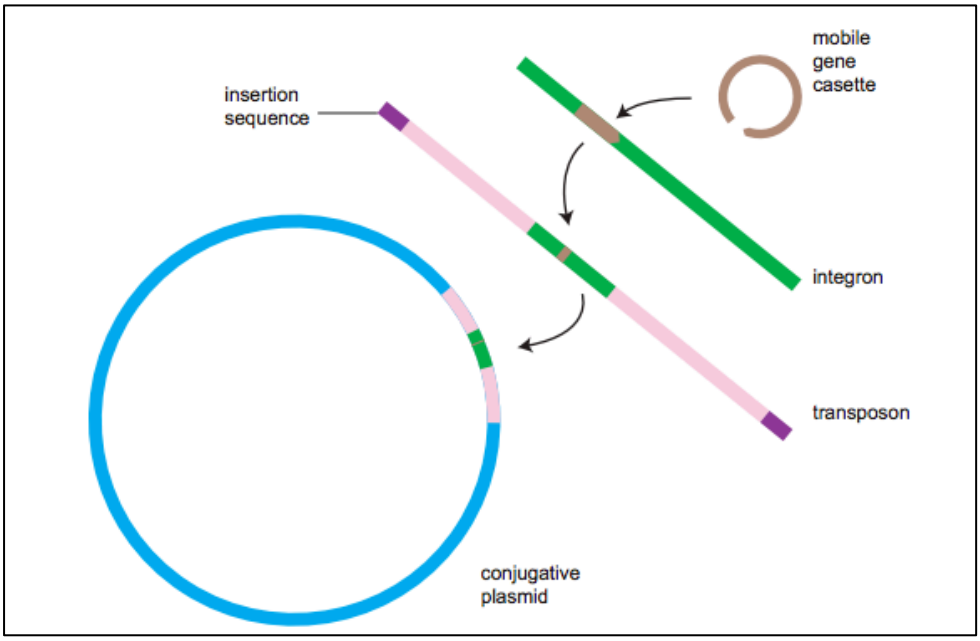


Figure 3: Mobile genetic elements compose of plasmids, transposons and integrons. Mobile gene cassettes typically carrying antimicrobial resistance genes are acquired and expressed in integrons by site specific recombination [53].

Flow of genetic information across different microbial communities depends on the proximity of potential recipient and donor bacteria, and the extent of LGT. The exchange and interaction of genes due to LGT has resulted in the formation of complex microbial communities within diverse environmental niches. Metagenomic studies on environmental samples found that a large amount of these samples contained MGE sequences in the form of phages, prophages and plasmids, in addition to chromosomes [54].

The evolution of prokaryotes may be altered by the environmental stress in which they live and the pool of compatible MGE in that environment. Many MGE can potentially move across all permissive host cells within an environment at favourable conditions. Movement of MGEs is

not only driven by selective pressure from antibiotic use or misuse, but also depends on the competition for resources and source of genetic information available in the relevant microbial communities. Over time this generates an accessory gene pool, that is generated through horizontal gene transfer and is commonly shared across different strains within certain species of bacteria such as those belonging to *Enterobacteriaceae*. These bacteria carrying these MGEs successfully persist due to the ability of specific genes such as AMR or virulence genes within the accessory genome to transfer and express in the pre-existing framework of host cells. The accessory gene pool is a massive resource for bacteria that provides flexibility and enhancing fitness and potentially pathogenicity and virulence [53] [54].

2.4.1 Transposable elements:

Transposons are associated with genome rearrangements and generate unique chromosome features within bacteria. Transposable elements (TE) are discrete DNA sequences capable of movement between replicons. Unlike plasmids, transposons cannot be mobilised between cells and are dependent on integration into independently replicating molecules. Transposons are flanked by inverted repeats at the boundaries of the mobile unit. Transposable elements can be categorised into two types based on their transposition intermediates. Class I TE or retrotransposons undergo replicative strand transfer mechanism of transposition. Instead of direct replication, reverse transcription occurs where RNA intermediate is converted to cDNA by reverse transcriptase that is integrated into the genome. Class II TE also called as DNA transposons that encode a transposase which facilitates DNA to either be excision and insertion non-replicative mechanism (**Figure 4**) [56].

The non-replicative mechanism is where the element is excised and inserted into a new target molecule. Characterised by terminal inverted repeats that encodes for transposases that binds near the inverted repeats and facilitates movement.An example of this is Tn*7* which carries the tnsABCDE operon. TnsA and TnsB together act as transposases that excise Tn*7* from its original site. The transposition generates a 5-bp target site duplication [57]. The replicative mechanism is a two stage replicative transposition event that involves DNA replication and generates a cointegrate that has a second copy of IS at the target site. This co-integration is resolved by resolvase that restores the donor molecule and has one copy in the target molecule [56]. Resolvase is a type of site-specific recombinase that resolves the replication intermediates, or the co-integrates, and facilitates separation of the two DNA molecules. It

excises a circular segment of DNA with two recombination (*res*) sites. An example of this is IS*26*, which is 820bp long with 14bp terminal inverted repeats and a *tnp26* gene encoding for transposase. It has been previously demonstrated that IS*26* can generate a co-integrate between the DNA molecule that it resides in and the target molecule. Within the co-integrates, the IS*26* is duplicated that results in two copies of IS*26* in direct orientation, one copy at each end of the two involved DNA sequences. The transposition results in an 8bp duplication at the target molecule [58].

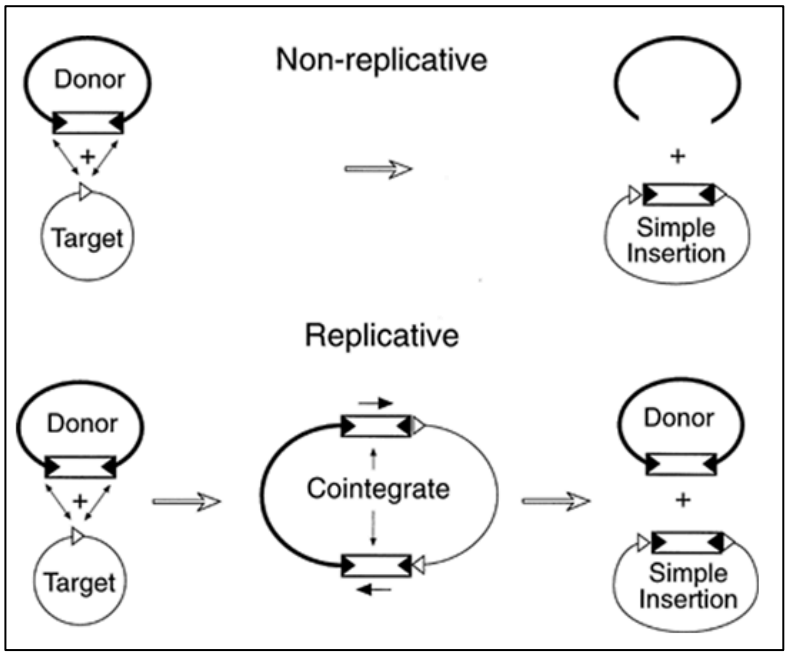


Figure 4: Non-replicative and replicative mechanism of transposition.

The size and genetic organisation of TEs is highly variable and range from large multiple drug resistant transposons to simple insertion sequences (IS). The most prevalent class of TE are the IS elements, that are typically acquired by plasmids. IS elements are the simplest form of transposable elements, consisting of a functional transposase enzyme (*tnp*) and a set of inverted repeats that is recognised by the transposase enzyme. An IS element does not carry accessory genes like antimicrobial resistance genes unlike transposons.

Large transposons are clinically important as they can mobilise a variety of antibiotic resistant genes. These transposons belonging to Tn*3* and Tn*5053* families are most commonly reported in Gram-negative bacteria. The Tn*3* family has two groups of transposons: Tn*21*-like and Tn*3*-like. They are both flanked by 38bp inverted repeats and consist of a transposase gene (*tnpA*),

resolvase gene (*tnpR*) and resolution site (*res*). These transposons mobilise by the replicative process (see above) and generate a 5bp direct repeat [59].

Two copies of an IS flanking resistance genes represent a composite transposon, however there are a number of atypical transposons that have been characterised, for example, IS*CR1* and IS*Ecp1*. These IS elements are unique as they have the ability to acquire and mobilise resistance genes in the absence of a second copy [57]. IS*CR1* elements are related to the IS*91* family of insertion elements that mobilise by rolling circle replication and do not generate direct repeats. ISCR1 has been previously associated with a number of antimicrobial resistance genes such as trimethoprim resistance (*dfrA18)*, sulphonamide resistance (*sul2)* as well as beta-lactamases ($bla_{CTX\text{-}M\text{-}2,\text{-}9}$ and $bla_{VEB\text{-}3}$) [60]. IS*Ecp1* is flanked by a third copy of a 14bp imperfect inverted repeat and creates a 5bp duplication on transposition. These are responsible for the mobilisation of a number of antimicrobial genes and extended spectrum beta-lactamases genes such as $bla_{CTX\text{-}M\text{-}15}$ that commonly appears in the dominant MDR *E. coli* ST131 clone [61].

2.4.2 Plasmids:

Plasmids are circular extra-chromosomal DNA entities that move via conjugation or transformation and are key elements of LGT. They are self-replicating via either the 'rolling circle' (RC) mechanism involving unidirectional replication of circular DNA, or by utilizing host machinery. They can remain as extra-chromosomal independently replicating entities or can integrate into the bacterial genome [62].

Plasmids are well known as mobile vectors for dissemination of virulence and antimicrobial resistance genes by lateral gene transfer. The movement of genes from the chromosome to these disruptive elements has been widely demonstrated. For instance, the pOLA52 plasmid from *E. coli* contains a type III fimbriae virulence gene. The fragment carrying the virulence and AMR genes was found to be captured by insertion of two IS*1* elements that was originally identified in the chromosome of *Klebsiella pneumoniae* [53]. This highlights that plasmids have the ability to not only acquire and mobilise genes but can change the genetic content of the bacterial chromosome.

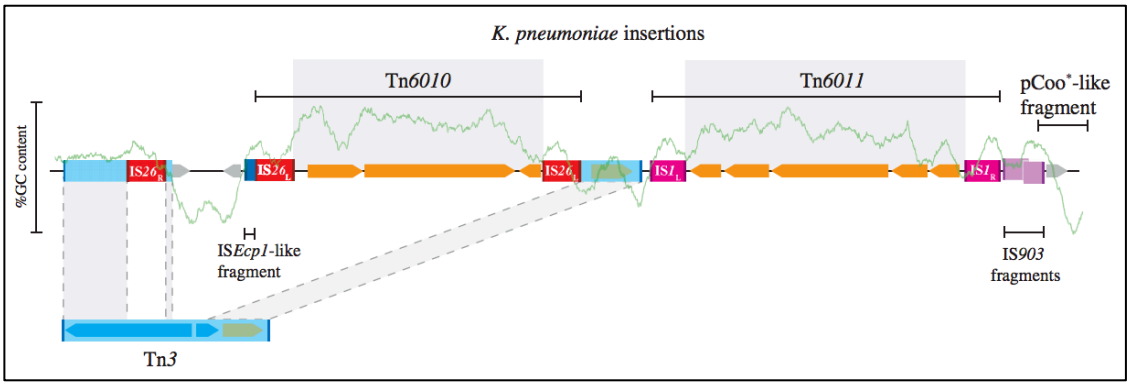


Figure 5: Structure of plasmid pOLA52 in *E. coli* that has two composite transposons (Tn*6010* and Tn*6011*) inserted that carries virulence gene associated with biofilm formation and IS elements (IS*1* and IS*26*). This structure has more than 99% homology to *K. pneumoniae* MGH 78578 [53].

As a result, examining MDR plasmids can provide a deeper insight into the evolution and dissemination of pathogens [53].

The core genes are typically conserved between broadly related plasmids, and are linked with plasmid specific functions like replication and mobility. Accessory gene appear to be found only in a subset of strains and move laterally between strains and form new trait combinations most commonly within the same species. Thus, different strains may have different set of accessory genes and a subset of all these unique genes within a species is a part of the pan-genome. The emergence of different clonal lineages with a species acquire and lose variable genes from other species in their local communities. The size of the accessory or variable genes that are distributed among different strains is an essential genetic resource for adaptation [64] [63]. The evolution of prokaryotes may be altered by the environmental stress in which they live and the pool of compatible MGE in that environment. Many MGE can potentially move across all permissive host cells within an environment at favourable conditions. Movement of MGEs is not only driven by selective pressure from antibiotic use or misuse, but also depends on the competition for resources and source of genetic information available in the relevant microbial communities. Over time this generates an accessory shared gene pool. Bacteria carrying these MGEs successfully persist due to the ability of specific genes such as AMR or virulence genes to transfer and express in the pre-existing framework of host cells and confer enhanced fitness for survival [53] [54].

A number of plasmids carry MDR regions and virulence factors that pose a grave threat to human health [65]. For example, a recent study of *E. coli* in drinking water in Tanzania found that >49% strains recovered carried a plasmid belonging to incompatibility group IncF, bearing antimicrobial resistance genes to ampillicin, cotrimoxazolem, tetracycline, gentamycin and amoxicillin/clavulanic acid and also harboured the globally pandemic extended spectrum beta-lactamase gene, $bla_{\text{CTX-M-15}}$ [66].

Plasmids have unique systems that ensure autonomous replication and also possess addiction systems that contribute to controlled copy number and stable inheritance during cell division. Plasmids are generally classified by typing schemes that exploit the backbone loci associated with plasmid replicon or plasmid mobility. Replicon incompatibility typing is traditionally used to classify plasmids into different incompatibility groups. Plasmid incompatibility occurs when two plasmids are unable to be stably inherited together or co-exist in the absence of external section [63].

Classifying plasmids provides a deeper insight into the epidemiology of plasmid mediated antibiotic resistance. Allelic profiles are assessed using 2 to 6 core loci, depending on the specific scheme. Plasmids are then assigned specific sub-types within the broader replicon type. This sub-typing of plasmids into different incompatibility groups has been commonly carried out using *in silico* plasmid multi-locus sequence typing (pMLST). There are six common replicon types circulating within *Enterobacteriaceae* that include IncF, IncHI1,-2, N, A/C [67].

IncF plasmid is the most common incompatibility group and has been isolated from a variety of sources. IncF plasmids are a diverse group that vary in size and replicon type. Notably, a number of studies worldwide have shown that IncF plasmids have the ability to integrate a range of genes that confer resistance to major classes of antimicrobials, namely, beta-lactams, aminoglycosides, quinolones, tetracycline, chloramphenicol and more recently carbapenems [68]. IncFII and IncFIB plasmids are known to be associated with the CTX-M type of extended spectrum beta-lactamases (ESBLs) such as $bla_{\text{CTX-M-14,-15,27}}$. It is concerning that these plasmids not only acquire antimicrobial resistance genes, but also specific virulence traits associated with bacterial fitness and survival found exclusively in IncF plasmids [68].

IncF plasmids can pose a major health threat as they are associated with pandemic clones such as MDR *E. coli* ST131 and ST95 that cause fatal infections in humans and food producing animals [64] [66]. A recent study demonstrated that the pandemics caused by MDR *E. coli* and *K. pneumoniae* are most likely caused by ST131 *E. coli* and ST258 *K. pneumonaie*. Both these globally disseminated clones harbour similar IncF plasmids carrying IncFIA and IncFIB replicons, suggesting that these promiscuous plasmids may have mobilised into different hosts causing outbreaks world-wide [69].

Notably, plasmids that carry antimicrobial resistance genes are described as R plasmids or R-factors. Some of these R-plasmids can transfer themselves via conjugation and promote host resistance to specific antimicrobials namely sulfonamides, tetracycline and streptomycin. Antimicrobial resistance can rapidly spread across different strains of bacteria as many of the R-factors contain F plasmids. R-factors can also be taken up via transformation and transduction. The genes present in them are called as resistance transfer factor (RTF) [58].

It is evident that plasmids are mobilising across different environments, however there is little understanding of the diversity of plasmids found in MDR *E. coli* causing fatal blood-stream infections in humans. It is crucial to understand the biology of MDR plasmids that are acquired or established under such selection pressures and their resistance and virulence genes [68]. Plasmid acquisition impacts persistence of bacterial clones under different environmental pressures. Studying these patterns would provide details on the evolution of pandemic or emerging clones [67-70] within *Enterobacterales* causing BSI.

2.4.3 Genomic islands (GIs):

Genomic islands (GIs) are chromosomal DNA segments that are generally found in strains with common ancestry and are often acquired by LGT [71]. Many GIs are mobile and carry genes that affect the pathobiology of the host organism and are involved in host fitness. GIs often carry clusters of antimicrobial resistance genes and are known to be responsible for bacterial genome diversification. Examples include: (i) satellite prophages, that are similar to defective prophages as they lack their own structural protein genes but are capable of encapsidation of DNA segments using other proteins from virions; and (ii) conjugative elements that reside in the host chromosome but are capable of excising DNA segments and transfer by conjugation [72].

Some genomic islands can be mobilised by plasmids that provide conjugative functions *in trans*. For example, IncA/C plasmids can confer conjugative functions to Salmonella Genomic Island 1 (SGI1). SGI1 harbours genes that encode for resistance to ampicillin, chloramphenicol, florphenicol, streptomycin, spectinomycin, tetracycline and sulphonamides. GIs carrying numbers of antimicrobial resistance genes may disseminate to other microbial communities, however this mechanism not yet clearly understood, as in many cases GIs lack known mobility related genes [73] [74].

2.4.4 Integrons:

Integrons play an important role in the spread of ARG's in both clinical and agricultural settings. Integrons have the ability to capture and express promoterless ARG cassettes which insert themselves via site specific recombination. The *intI* gene encodes an integrase protein and belongs to the tyrosine recombinase family. The integrase protein facilitates site specific recombination between the *attI* site of the core integron and the *attC* site of the gene cassette. Site specific recombination occurs between the *attC* site or 59 base element (of 57-141 base pairs) present in the gene cassette and the *attI* site present in the core integron **(Figure 2)**. The integron integrase protein specifically recognises the *attI* site in the 5'-CS. There are many different classes of integron however only classes 1, 2 and 3 are clinically significant. Class 1 integrons are the most prevalent integrons present in 40-70% of pathogenic Gram-negative bacteria carrying ARG clusters that disseminate in different bacterial populations [75].

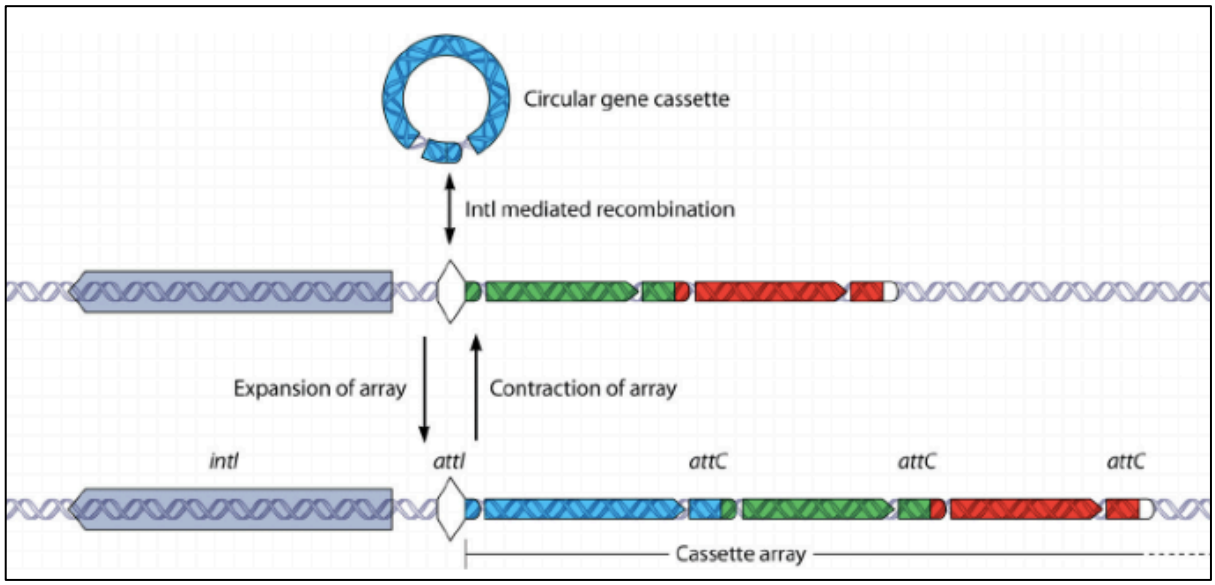


Figure 6: Structure of a core integrons and acquisition of gene casettes [75].

Structurally, class 1 integrons found in clinically important bacteria are bounded by inverted repeats, IR_i at the 5'-end and IR_t at the 3'-end and consist of a truncated transposition module (tni module) associated with a Tn*402* transposon present between the 3'-*CS* and the IR_t end. The 5'-*CS* contains the core integrase and the promoter (P_c) whereas the 3'-*CS* in most class 1 integrons includes sulfonamide resistance gene *sul1,* quaternary ammonium compound *qacEΔ1* resistant gene (Δ denotes qac gene truncation) and *orf5* gene [76]. Generally, multiple cassettes remain embedded in the core integron structure forming a cassette array, and results in the expression of all genes in the array [77].

Cassette arrays have the flexibility to rearrange the order of their gene cassettes and alter levels of expression based on the selective pressure in bacteria [77]. The successful capture, expression and maintenance of gene cassettes by clinical class 1 integrons suggests that these capture systems are not only able to disseminate but also persist under selective pressure [79].

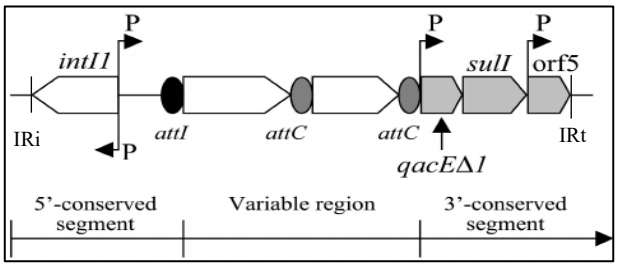


Figure 7: Typical structure of class 1 integrons prevalent in the clinically important bacteria [76].

In most class 1 integrons reported from clinical samples, the *tni* module of the ancestral Tn*402* transposon is truncated, rendering them defective. Class 1 integrons are hence commonly immobile. Plasmids, IS elements and transposons mobilise class 1 integrons found in the clinics. Recent studies have revealed the presence of class 1 integrons with complete functional Tn*402* modules and Tn*402* hybrid modules facilitating movement of integrons [79] [80].

As class 1 integrons primarily harbour cassettes with multiple AMR genes in arrays, they can serve as a proxy for identification of MDR regions in Gram-negative pathogen genomes, especially in members of *Enterobacterales* including *Escherichia coli*. Typically one to seven gene cassettes can incorporate into clinical class 1 integrons. Notably, there are a large number of gene cassettes encoding resistance genes [81] [82] and a majority of these are not only from clinical bacterial populations but also from food producing animals, soil and water bodies.

These unique recombination systems are important for persistence of bacterial populations, and are reported in different sources and different plasmids and transposons related to the widespread administration of antimicrobials [81]. Resistance gene cassettes are not only restricted to class 1 integron structures prevalent in clinical environments. For example, a recent study in South America on surface water and soil samples, identified gene cassettes encoding trimethoprim (*dfrA*), streptomycin and spectinomycin (*aadA*) resistance, but with less diversity. The genes included *dfrA17, dfrA1, dfrA12* and *aadA1, aadA5* and *sul1* associated with class 1 integrons that are also often found in clinical samples. The prevalence, evolution, and rapid dissemination of clinical class 1 integrons demonstrate bacterial evolution in action, driven by the selective pressure exerted by antimicrobial drugs [83].

The exact events that lead to the formation of clinical class 1 integrons are not known, however it is hypothesized that it started when a Tn*402*-like transposon captured the class 1 integron. After the introduction sulphonamides in the late 1930s, there was acquisition of the alternative dihydropteroate synthase gene, *sul1* gene by the Tn*402*-like integron. The end of the *qacE* gene was deleted and its *attC* site generated the 3'-*CS* end that is typically seen in clinical class 1 integrons. Further deletions to the Tn*402* element resulted in loss of transposition functions and led to the diversity in the 3'-*CS* end of clinical class 1 integrons [75]. Class 1 integrons have lost Tn*402* transposition functions, however the success of Tn*402*-like integrons may partly be due to the adaptive property of Tn*402* that is known to target and transpose into the *res* site of plasmids. The most well-known of these insertion events is the transposition of Tn*402* integron into a mobile element harbouring the mercury module, resulting in the In*2* integron bearing the Tn*21* transposon [75].

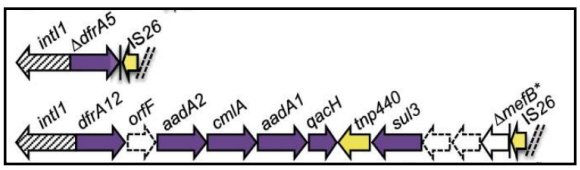


Figure 8: Schematic representation of clinical class 1 integrons associated with IS*26* and ARG cassettes [87].

Until recently, PCR targeting the two conserved regions to amplify genes in the variable region was one of the most common methods for surveillance of class 1 integrons [85]. However, the

3'-*CS* in class 1 integrons may be altered and, as a result, clinically important integrons may remain undetected by PCR. These class 1 integrons with such atypical structures are formed by recombination or interruption by insertion elements that skew primer binding sites. There are numerous reports of the presence of atypical structures of class 1 integron found in the clinics within complex ARG loci [84]. For instance, recent studies on ExPEC MDR strains demonstrate that the In*2* type of clinical class 1 integron is found to be interrupted by IS*26* associated Tn*6029* transposon, where IS*26* flanks this structure that carries ARGs, *sul2, strA/B* and bla-TEM-1B [86]. In another example found in porcine and clinical *E. coli* isolates, IS*26* altered the 3'-CS region by inserting itself next to the trimethoprim resistance encoding gene, *dfrA5* **(Figure 8)**. IS*26* deletions of *mefB* gene in *sul3* associated class 1 integrons result in formation of complex ARG loci [87-89]. Presence of a functional class 1 integron provides a scaffold for capturing and expressing emerging and novel antimicrobial resistance genes. However, other MGE such as IS*26* should not be underestimated as their presence can significantly impact the evolution and dissemination of MDR bacteria in human and food producing animals [90].

2.5 Fitness and survival of *E. coli* clones:

The acquisition of virulence-associated genes provides an evolutionary road map to pathogenicity in infections. Virulence factor's and bacterial fitness play an influential role in the spread of antibiotic resistant strains. If the virulence genetic determinant is present on the same plasmid or genetic element as the antibiotic resistant gene, then these genes may co-mobilise under selective pressure to form stable virulent clones.

The virulence potential in *E. coli* and other bacterial populations can be determined by presence of virulence factors such as fimbrae, adhesins, toxins, siderophores, capsules, hemolysins and invasins. These virulence factors help *E. coli* survive and colonise in the presence of host defense mechanisms and may produce host inflammatory responses [91].

Clinical pathologic outcomes are closely related to virulence gene combinations in *E. coli* [92]. Traditionally, pathogenic *E. coli* classification was dependent on detecting virulence associated genes to classify them as ExPEC, enterotoxigenic (ETEC), enteropathogenic (EPEC), enterohemorrhagic (EHEC), enteroinvasive (EIEC) and enteroaggressive (EaggEC) [93].

EXPEC strains exhibit a variety of extraintestinal virulence factors that facilitates them to colonize host mucosal surfaces, resist local and systemic host defence systems, provide important nutrients like iron and injure and invade the host and trigger a fatal inflammatory response. These strains typically produce virulence factors such as; 1) adhesins, that allows bacteria to bind to host cells, often encodes on the bacterial backbone DNA such as S fimbriae and Type 1 fimbriae; 2) polysacchride coating like lipopolysacchrides and capsules enhances the ability of bacteria to cause disease as it encapsulates it and protects it from the host; 3) toxins coat the bacterial surface and inhibit phagocytosis and surface receptors that bind to host cells; 4) siderophores or aerobactins act as iron scavenging chemicals as they chelate iron ions from the host environment; 5) invasins promote attachment and invasion of host cells; 6) serum resistance is the ability of bacteria to survive in the serum, a crucial virulence factor that is typically associated with blood-stream infections or urosepsis [95, 93].

A study by Johnson et al, demonstrated that a strain can be defined as an ExPEC if it carries at least two of the virulence factors such as P fimbria (*papA, papC*), specific adhesins (*afa* or *dfra*), type II capsule (*kpsMT* II) and aerobactin (*iucA* or *iutA*) [94]. Using this study, Santos et al investigated the presence of virulence factors in ExPEC and non-ExPEC strains causing bacteremia to reveal that 61% *E. coli* strains possessed intrinsic virulence as compared to 39% that had low virulence [92]. However, the *E. coli* genome was found to be highly diverse, and the emergence of hybrid strains made it hard to designate pathotypes based on specific virulence factors [95].

ExPEC strains most commonly cause bacteraemia [96] and typically carry multiple virulence associated genes that encode for adhesins, iron acquisition, toxins, protectins, biofilm production and gut colonisers. ExPEC strains possess virulence gene combinations that are distinctive compared to other intestinal infections [97]. Virulence associated genes play a role in infection by assisting bacteria to not only to invade host cells and tissues, but to also mediate iron availability, exert toxic effects on the host or provide protection against the host immune system [98]. Reports demonstrate that certain virulence associated genes such as *pap*, *malX, fyuA* and pylogroup B are associated with high mortality rates in ExPEC infections [99].

These virulence factors, along with the host response, induced epithelial cell dysfunction and increased cell membrane permeability, enable bacteria to enter the blood-stream [100]. Virulence-associated genes affect the occurrence and outcome of bacteraemia in humans. A

recent study showed that ST131 *E. coli* clones are typically associated with *Iha, hlyD, sat, Iut, fyuA, traT, OmpT* and *malX* genes compared to non-ST131 *E. coli* clones [101]. This suggests that the higher epidemiological success of ST131 strains may be due to the presence of specific virulence factors in these strains. It is important to note that specific virulence-associated genes may be linked to particular STs and infections.

High virulence potential is often found in strains from patients that are immunocompetent, as the bacteria requires additional virulence factors to overcome the hosts immune system [101]. Low virulence strains are typically associated with HA infections since more virulence factors are not needed when the host defence is already impaired. Compared to these infections, CO infections would be associated with higher virulence for persistence and dissemination of these infections in the community. However, a recent study in Demark demonstrated that two virulence-associated genes known to be prevalent in BSI, *kpSM* II and *hylD*, were found in HO-BSI rather than CO-BSI [99]. The discrepancies in these studies demonstrate that a deeper understanding of virulence associated genes in *E. coli* BSI is necessary, and that this information would be useful to identify the localisation of these strains within various niches and hosts, facilitating further epidemiological studies [92].

2.6 *E. coli* sequence types and phylogeny:

E. coli typically colonize the gut in humans and animals and can cause infection in intestinal or extra-intestinal sites. A number of tools have been developed for rapid identification of *E. coli* clones or clonal complexes, serotype and phylogroups. These tools allow the ecological niche, disease severity to cause disease, pathogenicity and evolutionary origins to be determined [102].

Conventionally, classification of *E. coli* strains was based on antigenic determinants on the cell surface detected by presence of O (Somatic), K (capsular polysaccharide) and H (flagellar) antigens. Serotypes O4/O6/O75/O25 and H4 were found in more than 50% of ExPEC strains [103]. However, one of the major disadvantages of serotyping is that it has a poor discriminatory power compared to new developed *in silico* tools or schemes in microbial epidemiology [105] [106]. The popular tools used for studying population structures of *E. coli* include multi-locus sequence typing (MLST), phylogrouping, and more recently, whole genome sequencing. Three MLST databases have been established for *E. coli* MLST. The *Ec*MLST database at Michigen State University focuses on enteropathogenic *E. coli*, while the

database hosted at Warwick Medical School and at the Pasteur Institute have no specific focus [104].

The rationale for the specific choice of genes for the different MLST schemes remains unclear, however, all use housekeeping genes. Currently there are three different MLST schemes for E. coli. 1) The Pasteur Institute scheme was based on 11 housekeeping genes 2) the Acthman scheme that is hosted by Warwick Medical School, UK, uses 7 housekeeping genes, on the basis of low lateral gene transfer, therefore providing higher ability to distinguish between different STs. Notably, across all the schemes, the Acthman scheme genes had the lowest nucleotide diversity and is one the most commonly used MLST scheme [107] [108]. However, although MLST provides sequence type (ST) information, it does not directly give phylogenetic information.

The phylogenetic grouping or phylogrouping method that is typically used for *E. coli* strains is the 'Clermont method'. This is based on two genes (*chuA* and *yjaA*) and one DNA fragment (TspE4.C2) that divided *E. coli* into A, B1, B2, D phylogroups [108]. However, with the expansion of the MLST database and use of WGS, there are now more than four types of phylogroups. Addition of the *arpA* gene allows the classification of *E. coli* into seven phylogroups, A, B1, B2, C, D, E and F [109].

E. coli has a number of common, clinically important lineages, such as clonal group A corresponding to ST68 as well as B2 lineages ST131, ST73 and ST95. *E. coli* ST131 is a globally disseminated MDR clone seen across diverse extra-intestinal infections that includes UTI and BSI. These clones typically harbour resistance genes against a variety of antibiotics that include trimethoprim-sulfamethoxazole, fluroquinolones, 3rd generation cephalosporins, amoxicillin and clavulanic acid. In addition, these strains are reported to carry a number of virulence genes and are known to be associated with pandemic extended spectrum beta-lactamase (ESBL) genes [110] [111].

From the global perspective, a recent survey of 169 studies indicated that the proportion of ExPEC ST131 lineages in each geographic region was higher compared to other major STs. ST131 clones were most commonly found in Asia, South America, Middle-East and Europe. Lower percentage of ST131 and others STs was observed in Africa. This may be due to small sample collection or MLST reference database limitation. This survey also indicated that from

the studies on ExPEC strains world-wide, twenty major STs were identified, of which ST131 was found in more than 90% of studies. The other major ST identified world-wide, included; ST69, ST10, ST95, ST405, ST ST38, ST95, ST648, ST73, ST410, ST393, ST354, ST12, ST127, ST167, ST58, ST617, ST88, ST23, ST117, and ST1193. Notably, ST95 and ST10 were found in more than 50% of the overall studies. The maximum of 84 STs was identified in one of the studies while other studies identified few STs [55].

Over the past decade, the rise in fluroquinolone resistance is due to rapid emergence of ST131 sub-types H30 and H30-Rx that have the tendency to encode for ESBL gene $bla_{CTX\text{-}M\text{-}15}$. A study in Canada on BSI demonstrated that the ST131 clone produce ESBL's such as $bla_{CTX\text{-}M\text{-}15}$ and $bla_{CTX\text{-}M\text{-}14}$ in community acquired *E. coli* bacteremia. A study in USA demonstrated that ST95 was the second most common ST in *E. coli* BSI. Reports in Europe, Japan, USA and Australia demonstrated that ST405 was one of the predominant STs in ESBL producing ExPEC pathogens [117]

Other increasingly reported MDR *E. coli* STs included ST10, ST38, ST117, ST1193 and ST393 [91] [112-114]. A recent epidemiological study in Saudi Arabia on uropathogenic *E. coli* isolates revealed that the common lineages associated with antibiotic resistance were ST131 (17%), ST73 (11.4%), ST38 and ST69 (7.4%), ST10 (6.4%), ST127 (5.9%), ST95 (5.4%), ST12 (3.5%), ST998 (3.5%) and ST405 (3%). Notably, 60% of ST131 isolates harboured $bla_{CTX\text{-}M\text{-}14}$ or $bla_{CTX\text{-}M\text{-}15}$ and 66% of ST405 isolates harboured $bla_{CTX\text{-}M\text{-}15}$ [115]. This demonstrates that diverse multi-drug resistant *E. coli* clones are emerging globally.

2.7 Conclusion:

Prior to this study, no systematic genomic wide surveillance of *E. coli* from community onset and hospital onset BSI had been performed in Australia. While previous studies have reported targeted virulence and resistance gene profiles in *E. coli* associated with BSI [37], there are no studies focused on identifying the relative abundance of resistance and virulence genes between hospital onset and community onset infections, or associations of resistance and virulence genes with mobile scaffolds that could lead to the rapid evolution of epidemic clones.

As MGE in *E. coli* and other Gram-negative pathogens are known to carry a cargo of niche-adaptive genes for bacterial survival, and provide an evolutionary advantage leading to fatal

infections like BSI, understanding the mechanisms of acquiring these genes will potentially provide attractive targets for new antibiotics. For instance, identified targets can be selected for genome editing in bacteria to control the acquisition and spread of MDR and consequently reduce the burden of infectious disease [51].

This pilot study investigates the movement of MGEs that carry AMR and virulence-associated genes in HO and CO-BSI *E. coli*. We identify specific MGE that are associated with the movement of specific resistance gene clusters and, for the first time, characterise novel and dominant sequence types of MDR *E. coli* BSI from hospital and community settings within Australia.

2.8 References:

1. Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. Microbiology and molecular biology reviews : MMBR, 74(3), 417–433. doi:10.1128/MMBR.00016-10
2. Ventola C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. P & T : a peer-reviewed journal for formulary management, 40(4), 277–283.
3. Graham, J. P., Eisenberg, J., Trueba, G., Zhang, L., & Johnson, T. J. (2017). Small-Scale Food Animal Production and Antimicrobial Resistance: Mountain, Molehill, or Something in-between?. Environmental health perspectives, 125(10), 104501. doi:10.1289/EHP2116
4. Taneja N, Sharma M. Antimicrobial resistance in the environment: The Indian scenario. Indian J Med Res 2019;149:119-28.
5. Thorpe KE, Joski P, Johnston KJ. Antibiotic-Resistant Infection Treatment Costs Have Doubled Since 2002, Now Exceeding $2 Billion Annually. Health Aff (Millwood). 2018 Apr;37(4):662-669. doi: 10.1377/hlthaff.2017.1153. Epub 2018 Mar 21. PubMed PMID: 29561692.
6. O'Neill J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. London: Wellcome Trust, UK Government2016.
7. Luyt CE, Bréchot N, Trouillet JL, Chastre J. Antibiotic stewardship in the intensive care unit. Crit Care. 2014 Aug 13;18(5):480. doi: 10.1186/s13054-014-0480-6. Review. PubMed PMID: 25405992; PubMed Central PMCID: PMC4281952.
8. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2013.
9. Australian Commission on Safety and Quality in Health Care. AURA 2016: first Australian report on antimicrobial use and resistance in human health. Sydney: ACSQHC; 2016.
10. Shaban R. Z., Cruickshank M., Christiansen K., The Antimicrobial Resistance Committee (2013). National Surveillance and Reporting of Antimicrobial Resistance and Antimicrobial Usage for Human Health in Australia.
11. Smalla K , Simonet P, Tiedje Jet al. Editorial: Special section of FEMS Microbiology Ecology on the environmental dimension of antibiotic resistance. FEMS Microbiol Ecol 2016;92:fiw172.

12. Lundborg, C. S., & Tamhankar, A. J. (2017). Antibiotic residues in the environment of South East Asia. BMJ (Clinical research ed.), 358, j2440. doi:10.1136/bmj.j244
13. Fick J, Söderström H, Lindberg RH, Phan C, Tysklind M, Larsson DG. Contamination of surface, ground, and drinking water from pharmaceutical production. Environ Toxicol Chem. 2009 Dec;28(12):2522-7. doi: 10.1897/09-073.1. Epub 2009 May 18. PubMed PMID: 19449981.
14. Bartlett JG, Gilbert DN, Spellberg B. Seven ways to preserve the miracle of antibiotics. Clin Infect Dis. 2013 May;56(10):1445-50. doi: 10.1093/cid/cit070. Epub 2013 Feb 12. PubMed PMID: 23403172.
15. Phillips I, Casewell M, Cox T, De Groot B, Friis C, Jones R, Nightingale C, Preston R, Waddell J. Does the use of antibiotics in food animals pose a risk to human health? A critical review of published data. J Antimicrob Chemother. 2004 Jan;53(1):28-52. doi: 10.1093/jac/dkg483. Epub 2003 Dec 4. Review. PubMed PMID: 14657094.
16. McEwen SA, Collignon PJ. Antimicrobial Resistance: a One Health Perspective. Microbiol Spectr. 2018 Mar;6(2). doi: 10.1128/microbiolspec.ARBA-0009-2017. Review. PubMed PMID: 29600770.
17. Nicholson, A., Tennant, I., White, L., Thoms-Rodriguez, C. A., Cook, L., Johnson, S., … Richards, L. (2018). A national survey of the knowledge, attitudes and prescribing practices of doctors regarding antibiotic resistance in a Caribbean country. Antimicrobial resistance and infection control, 7, 23. doi:10.1186/s13756-018-0315-3
18. Polat, G., Ugan, R. A., Cadirci, E., & Halici, Z. (2017). Sepsis and Septic Shock: Current Treatment Strategies and New Approaches. The Eurasian journal of medicine, 49(1), 53–58. doi:10.5152/eurasianjmed.2017.17062
19. Viscoli C. (2016). Bloodstream Infections: The peak of the iceberg. Virulence, 7(3), 248–251. doi:10.1080/21505594.2016.1152440
20. Heldens M, Schout M, Hammond NE, Bass F, Delaney A, et al. (2018) Sepsis incidence and mortality are underestimated in Australian intensive care unit administrative data. Med J Aust 209(6): 255-260.
21. de Kraker MEA, Jarlier V, Monen JCM, Heuer OE, van de Sande N, Grundmann H.. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. Clin Microbiol Infect 2013; 19(9):860-8; PMID:23039210; http://dx.doi.org/10.1111/1469-0691.12028
22. Rodriguez-Bano J, Picon E, Gijon P, et al. Risk factors and prognosis of nosocomial bloodstream infections caused by extended-spectrum-beta-lactamase-producing *Escherichia coli*. J Clin Microbiol. 2010;48:1726–1731
23. Livermore D. M. (2012). Current epidemiology and growing resistance of gram-negative pathogens. The Korean journal of internal medicine, 27(2), 128–142. doi:10.3904/kjim.2012.27.2.128
24. Kallen AJ, Hidron AI, Patel J, Srinivasan A. Multidrug resistance among gram-negative pathogens that caused healthcare-associated infections reported to the National Healthcare Safety Network, 2006-2008. Infect Control Hosp Epidemiol. 2010 May;31(5):528-31. doi: 10.1086/652152. PubMed PMID: 20334552.
25. Joint Expert Technical Advisol)' Committee on Antibiotic Resistance (JETCAR), Commonwealth Department of Health and Aged Care, Commonwealth Department of Agriculture, Fisheries and ForestryAustralia. The Use of Antibiotics in Food-Producing Animals: Antibiotic-Resistant Bacteria in Animals and Humans. Commonwealth of Australia, 1999.

26. Hollenbeak CS. The cost of catheter-related bloodstream infections: implications for the value of prevention. J Infus Nurs. 2011 Sep-Oct;34(5):309-13. doi: 10.1097/NAN.0b013e3182285e43. PubMed PMID: 21915004.
27. Thaden, J. T., Li, Y., Ruffin, F., Maskarinec, S. A., Hill-Rorie, J. M., Wanda, L. C., … Fowler, V. G., Jr (2017). Increased Costs Associated with Bloodstream Infections Caused by Multidrug-Resistant Gram-Negative Bacteria Are Due Primarily to Patients with Hospital-Acquired Infections. Antimicrobial agents and chemotherapy, 61(3), e01709-16. doi:10.1128/AAC.01709-16
28. Friedman ND, Kaye KS, Stout JE, McGarry SA, Trivette SL, Briggs JP, Lamm W, Clark C, MacFarquhar J, Walton AL, Reller LB, Sexton DJ: Health care-associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. Ann Intern Med. 2002, 137: 791-797. 10.7326/0003-4819-137-10-200211190-00007.
29. Nimmo GR, Bell JM, Collignon PJ. Surveillance by the Australian Group for Antimicrobial Resistance (AGAR). Commun Dis Intell 2013;27 Suppl:S47–S54.
30. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2016. Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Stockholm: ECDC; 2016.
31. Peleg, A. Y., & Hooper, D. C. (2010). Hospital-acquired infections due to gram-negative bacteria. The New England journal of medicine, 362(19), 1804–1813. doi:10.1056/NEJMra0904124
32. Prowle, J. R., Echeverri, J. E., Ligabo, E. V., Sherry, N., Taori, G. C., Crozier, T. M., … Bellomo, R. (2011). Acquired bloodstream infection in the intensive care unit: incidence and attributable mortality. Critical care (London, England), 15(2), R100. doi:10.1186/cc10114
33. Tumbarello M, Viale P, Viscoli C, Trecarichi EM, Tumietto F, Marchese A, Spanu T, Ambretti S, Ginocchio F, Cristini F, Losito AR, Tedeschi S, Cauda R, Bassetti M. Predictors of mortality in bloodstream infections caused by *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae*: importance of combination therapy. Clin Infect Dis. 2012 Oct;55(7):943-50. doi: 10.1093/cid/cis588. Epub 2012 Jul 2. PubMed PMID: 22752516.
34. Buetti N, Marschall J, Atkinson A, Kronenberg A. National Bloodstream Infection Surveillance in Switzerland 2008-2014: Different Patterns and Trends for University and Community Hospitals. Infect Control Hosp Epidemiol. 2016 Sep;37(9):1060-7. doi: 10.1017/ice.2016.137. Epub 2016 Jun 28. PubMed PMID: 27350313.
35. Wilson J. (2018). Applying Pareto analysis to reducing *Escherichia coli* bloodstream infections. Journal of infection prevention, 19(5), 208–210. doi:10.1177/1757177418795298
36. Valles, J., J. Rello, A. Ochagavia, J. Garnacho, and M. A. Alcala. 2003. Community-acquired bloodstream infection in critically ill adult patients: impact of shock and inappropriate antibiotic therapy on survival. Chest 123:1615–1624.
37. Akova M. (2016). Epidemiology of antimicrobial resistance in bloodstream infections. Virulence, 7(3), 252–266. doi:10.1080/21505594.2016.1159366
38. Ahmed, D., Nahid, M. A., Sami, A. B., Halim, F., Akter, N., Sadique, T., … Rahman, M. M. (2017). Bacterial etiology of bloodstream infections and antimicrobial resistance in Dhaka, Bangladesh, 2005-2014. Antimicrobial resistance and infection control, 6, 2. doi:10.1186/s13756-016-0162
39. SA Health Data and Reporting Services. SA Healthcare-associated Infection Surveillance Program: Government of South Australia, 2015

40. Australian Commission on Safety and Quality in Health Care. AURA 2016: first Australian report on antimicrobial use and resistance in human health. Sydney: ACSQHC; 2016
41. Turnidge J, Gottlieb T, Mitchell D, Pearson J, Bell J, for the Australian Group for Antimicrobial Resistance. Gram-negative Survey 2011 Antimicrobial Susceptibility Report. 2011 Adelaide. Available from: http://www.agargroup.org/files/AGAR%20GNB08%20Report%20FINAL.pdf
42. Coombs G, Daley D, Gottlieb T, Turnidge JD, Bell JM. Australian Group on Antimicrobial Resistance: Sepsis Outcome Programs 2015 report. Sydney: ACSQHC; 2017
43. Annual epidemiological commentary: mandatory MRSA, MSSA and E. coli bacteraemia and C. difficile infection data, 2014/15.
44. Yahav, D., Eliakim-Raz, N., Leibovici, L., & Paul, M. (2016). Bloodstream infections in older patients. Virulence, 7(3), 341–352. doi:10.1080/21505594.2015.1132142
45. Aung AK, Skinner MJ, Lee FJ, Cheng AC. Changing epidemiology of bloodstream infection pathogens over time in adult non-specialty patients at an Australian tertiary hospital. Commun Dis Intell Q Rep. 2012 Dec 31;36(4):E333-41. PubMed PMID: 23330707.
46. Furuya, E.Y. & Lowy, F.D., 2006. Antimicrobial-resistant bacteria in the community setting. Nat Rev Micro, 4(1), pp.36–45. Available at: http://dx.doi.org/10.1038/nrmicro1325.
47. Thomas, C.M. & Nielsen, K.M., 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat.Rev.Microbiol., 3(1740–1526 (Print)), pp.711–721. Available at: C:\Karsten\PDFs\Bakteriologie-PDFs\Bakt-2005\Thomas - Nielsen-Mechanisms of and barriers to horizontal gene transfer between bacteria.pdf.
48. Ochman, H., Lawrence, J.G. & Groisman, E. a, 2000b. Lateral gene transfer and the nature of bacterial innovation. Nature, 405(6784), pp.299–304.
49. Melton-Celsa A, Mohawk K, Teel L, O'Brien A. Pathogenesis of Shiga-toxin producing *Escherichia coli.* Curr Top Microbiol Immunol. 2012;357:67-103. doi: 10.1007/82_2011_176. Review. PubMed PMID: 21915773.
50. Popa, O. & Dagan, T., 2011. Trends and barriers to lateral gene transfer in prokaryotes. Current Opinion in Microbiology, 14(5), pp.615–623. Available at: http://dx.doi.org/10.1016/j.mib.2011.07.027.
51. Jiang, X., 2017. Comprehensive analysis of mobile genetic elements in the gut microbiome reveals a phylum-level niche-adaptive gene pool. Preprint at https://www.biorxiv.org/content/early/2017/11/06/214213.
52. Frost, L.S. et al., 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Micro, 3(9), pp.722–732. Available at: http://dx.doi.org/10.1038/nrmicro1235.
53. Norman, A., Hansen, L. H., & Sørensen, S. J. (2009). Conjugative plasmids: vessels of the communal gene pool. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 364(1527), 2275–2289. doi:10.1098/rstb.2009.0037
54. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004 Apr 2;304(5667):66-74. doi: 10.1126/science.1093857. Epub 2004 Mar 4. PubMed PMID: 15001713.

55. Goldenfeld N.& Woese C.. 2007Biology's next revolution. Nature 445, 369. (doi:10.1038/445369a).
56. Hallet, B. & Sherratt, D.J., 1997. Transposition and site-specific recombination: Adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS Microbiology Reviews, 21(2), pp.157–178.
57. Partridge, S.R., 2011. Analysis of antibiotic resistance regions in Gram-negative bacteria. FEMS Microbiology Reviews, 35(5), pp.820–855.
58. Harmer, C.J., Moran, R.A. & Hall, R.M., 2014. Movement of IS*26*-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS*26* and preferentially inserts adjacent to another IS26. mBio, 5(5), pp.1–9.
59. Liebert, C. A., Hall, R. M., & Summers, A. O. (1999). Transposon Tn*21*, flagship of the floating genome. Microbiology and molecular biology reviews : MMBR, 63(3), 507–522.
60. Toleman, M.A., Bennett, P.M. & Walsh, T.R., 2006. Common regions e.g. orf513 and antibiotic resistance: IS*91*-like elements evolving complex class 1 integrons. Journal of Antimicrobial Chemotherapy, 58(1), pp.1–6.
61. Rossolini G. M., D'Andrea M. M., Mugnaioli C. (2008). The spread of CTX-M-type extended-spectrum β-lactamases. Clin. Microbiol. Infect. 14, 33–4110.1111/j.1469-0691.2007.01867
62. Meyer, R., 2009. Replication and conjugative mobilization of Broad Host-range IncQ plasmids. Plasmid, 62(2), pp.57–70.
63. Orlek, A., Stoesser, N., Anjum, M. F., Doumith, M., Ellington, M. J., Peto, T., … Sheppard, A. E. (2017). Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of Antibiotic Resistance Epidemiology. Frontiers in microbiology, 8, 182. doi:10.3389/fmicb.2017.00182
64. Phan M. D., Kidgell C., Nair S., Holt K. E., Turner A. K., Hinds J., et al. (2009). Variation in Salmonella enterica serovar typhi IncHI1 plasmids during the global spread of resistant typhoid fever. Antimicrob. Agents Chemother. 53 716–727. 10.1128/AAC.00645-08
65. Garcillán-Barcia M. P., Alvarado A., De la Cruz F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. FEMS Microbiol. Rev. 35 936–956. 10.1111/j.1574-6976.2011.00291.x
66. Lyimo, B., Buza, J., Subbiah, M., Temba, S., Kipasika, H., Smith, W., & Call, D. R. (2016). IncF Plasmids Are Commonly Carried by Antibiotic Resistant *Escherichia coli* Isolated from Drinking Water Sources in Northern Tanzania. International journal of microbiology, 2016, 3103672. doi:10.1155/2016/3103672
67. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.
68. Yang, Q. E., Sun, J., Li, L., Deng, H., Liu, B. T., Fang, L. X., … Liu, Y. H. (2015). IncF plasmid diversity in multi-drug resistant *Escherichia coli* strains from animals in China. Frontiers in microbiology, 6, 964. doi:10.3389/fmicb.2015.00964
69. Mathers, A. J., Peirano, G., and Pitout, J. D. (2015). The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant in *Enterobacteriaceae*. Clin. Microbiol. Rev. 28, 565–591. doi: 10.1128/CMR.00116-14
70. Moran RA, Hall RM. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. Microb Drug Resist 2017.
71. Juhas, M. et al., 2009. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. FEMS Microbiology Reviews, 33(2), pp.376–393.

72. Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. Annu. Rev. Genet. 49, 577–601. doi: 10.1146/annurev-genet-112414-055018
73. Carraro, Nicolas et al. "Mobilizable genomic islands, different strategies for the dissemination of multidrug resistance and other adaptive traits." Mobile genetic elements vol. 7,2 1-6. 12 Apr. 2017, doi:10.1080/2159256X.2017.1304193
74. Reid CJ, Roy Chowdhury P, Djordjevic SP. Tn*6026* and Tn*6029* are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant *Enterobacteriaceae*. Plasmid 2015;80:127-137.
75. Gillings MR. Integrons: past, present, and future. Microbiol Mol Biol Rev 2014;78(2):257-277.
76. Drouin, F., Mélançon, J. & Roy, P.H., 2002. The intI-like tyrosine recombinase of Shewanella oneidensis is active as an integron integrase. Journal of Bacteriology, 184(6), pp.1811–1815.
77. Stokes, H.W. & Hall, R.M., 1989. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Molecular Microbiology, 3(12), pp.1669–1683. Available at: http://doi.wiley.com/10.1111/j.1365-2958.1989.tb00153.x [Accessed September 28, 2016
78. Didelot, X. et al., 2012. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. BMC Genomics, 13(1), pp.1–15. Available at: http://dx.doi.org/10.1186/1471-2164-13-256.
79. Martinez, E., Marquez, C., Ingold, A., Merlino, J., Djordjevic, S. P., Stokes, H. W., & Chowdhury, P. R. (2012). Diverse mobilized class 1 integrons are common in the chromosomes of pathogenic *Pseudomonas aeruginosa* clinical isolates. Antimicrobial agents and chemotherapy, 56(4), 2169–2172. doi:10.1128/AAC.06048-11
80. Betteridge T, Partridge SR, Iredell JR, Stokes HW. Genetic context and structural diversity of class 1 integrons from human commensal bacteria in a hospital intensive care unit. Antimicrob Agents Chemother. 2011 Aug;55(8):3939-43. doi: 10.1128/AAC.01831-10. Epub 2011 May 31. PubMed PMID: 21628540; PubMed Central PMCID: PMC3147655.
81. Partridge SR, Tsafnat G, Coiera E, Iredell JR. Gene cassettes and cassette arrays in mobile resistance integrons. FEMS Microbiol Rev 2009;33(4):757-784.
82. Khoramrooz, S. S., Sharifi, A., Yazdanpanah, M., Malek Hosseini, S. A., Emaneini, M., Gharibpour, F., … Khosravani, S. A. (2016). High Frequency of Class 1 Integrons in *Escherichia coli* Isolated From Patients With Urinary Tract Infections in Yasuj, Iran. Iranian Red Crescent medical journal, 18(1), e26399. doi:10.5812/ircmj.26399
83. Canal, N., Meneghetti, K. L., de Almeida, C. P., da Rosa Bastos, M., Otton, L. M., & Corção, G. (2016). Characterization of the variable region in the class 1 integron of antimicrobial-resistant *Escherichia coli* isolated from surface water. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology], 47(2), 337–344. doi:10.1016/j.bjm.2016.01.015
84. Venturini C, Hassan KA, Roy Chowdhury P, Paulsen IT, Walker MJ et al. Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS26-related molecular signature isolated from different *Escherichia coli* pathotypes from different hosts. PLoS One 2013;8(11):e78862
85. Cocchi S, Grasselli E, Gutacker M, Benagli C, Convert M & Piffaretti JC (2007) Distribution and characterization of integrons in *Escherichia coli* strains of animal and human origin. FEMS Immunol Med Mic 50: 126–132

86. Roy Chowdhury, P., Charles, I.G., and Djordjevic, S.P. (2015). A role for Tn*6029* in the evolution of the complex antibiotic resistance gene loci in genomic island 3 in enteroaggregative hemorrhagic *Escherichia coli* O104:H4. PLoS One 10, e0115781. doi: 10.1371/journal.pone.0115781.
87. Reid, C. J., Wyrsch, E. R., Roy Chowdhury, P., Zingali, T., Liu, M., Darling, A. E., … Djordjevic, S. P. (2017). Porcine commensal *Escherichia coli*: a reservoir for class 1 integrons associated with IS*26*. Microbial genomics, 3(12), e000143. doi:10.1099/mgen.0.000143

88. Cummins M. L., Roy Chowdhury P., Marenda M. S., Browning G. F., Djordjevic S. P. (2019). Salmonella genomic island 1B variant found in a sequence type 117 avian pathogenic *Escherichia coli* isolate. mSphere 4:e169-19. 10.1128/mSphere.00169-119
89. Moran RA, Holt KE, Hall RM. pCERC3 from a commensal ST95 *Escherichia coli*: A ColV virulence-multiresistance plasmid carrying a sul3-associated class 1 integron. Plasmid. 2016 Mar-May;84-85:11-9. doi: 10.1016/j.plasmid.2016.02.002. Epub 2016 Feb 18. PubMed PMID: 26855083.
90. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect Inst Pasteur 2003; 5:449–56.
91. de Toro M, Sáenz Y, Cercenado E, Rojo-Bezares B, García-Campello M, Undabeitia E, Torres C. Genetic characterization of the mechanisms of resistance to amoxicillin/clavulanate and third-generation cephalosporins in *Salmonella enterica* from three Spanish hospitals. Int Microbiol. 2011 Sep;14(3):173-81. doi: 10.2436/20.1501.01.146. PubMed PMID: 22101415.
92. a Silva GJ, Mendonça N. Association between antimicrobial resistance and virulence in *Escherichia coli*. Virulence. 2012 Jan-Feb;3(1):18-28. doi: 10.4161/viru.3.1.18382. Epub 2012 Jan 1. Review. PubMed PMID: 22286707.
93. Pires-dos-Santos T, Bisgaard M, Christensen H. Genetic diversity and virulence profiles of *Escherichia coli* causing salpingitis and peritonitis in broiler breeders. Vet Microbiol. 2013 Mar 23;162(2-4):873-80. doi: 10.1016/j.vetmic.2012.11.008. Epub 2012 Nov 9. PubMed PMID: 23201240.
94. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol 2004;2(2):123-140.
95. Johnson JR. Microbial virulence determinants and the pathogenesis of urinary tract infection. Infect Dis Clin North Am. 2003 Jun;17(2):261-78, viii. doi: 10.1016/s0891-5520(03)00027-8. Review. PubMed PMID: 12848470.
96. Lindstedt BA, Finton MD, Porcellato D, Brandal LT. High frequency of hybrid *Escherichia coli* strains with combined Intestinal Pathogenic Escherichia coli (IPEC) and Extraintestinal Pathogenic Escherichia coli (ExPEC) virulence factors isolated from human faecal samples. BMC Infect Dis. 2018 Nov 1;18(1):544. doi: 10.1186/s12879-018-3449-2. PubMed PMID: 30497396; PubMed Central PMCID: PMC6267907.
97. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect Inst Pasteur 2003; 5:449–56.
98. Chapman, T. A., Wu, X. Y., Barchia, I., Bettelheim, K. A., Driesen, S., Trott, D., … Chin, J. J. (2006). Comparison of virulence gene profiles of *Escherichia coli* strains isolated from healthy and diarrheic swine. Applied and environmental microbiology, 72(7), 4782–4795. doi:10.1128/AEM.02885-05

99. Parvez, Sorwer & Rahman, Dolilur. (2018). Virulence Factors of Uropathogenic *E. coli*. 10.5772/intechopen.79557.
100. Skjøt-Rasmussen, Line & Ejrnæs, Karen & Lundgren, Bettina & Hammerum, Anette & Frimodt-Møller, Niels. (2012). Virulence factors and phylogenetic grouping of *Escherichia coli* isolates from patients with bacteraemia of urinary tract origin relate to sex and hospital- vs. Community-acquired origin. International journal of medical microbiology : IJMM. 302. 129-34. 10.1016/j.ijmm.2012.03.002.

101. Krawczyk B, Śledzińska A, Szemiako K, Samet A, Nowicki B, Kur J. Characterisation of *Escherichia coli* isolates from the blood of haematological adult patients with bacteraemia: translocation from gut to blood requires the cooperation of multiple virulence factors. Eur J Clin Microbiol Infect Dis. 2015 Jun;34(6):1135-43. doi: 10.1007/s10096-015-2331-z. Epub 2015 Feb 6. PubMed PMID: 25655758; PubMed Central PMCID: PMC4426128.
102. Hung WT, Cheng MF, Tseng FC, Chen YS, Shin-Jung Lee S, Chang TH, Lin HH, Hung CH, Wang JL. Bloodstream infection with extended-spectrum beta-lactamase-producing *Escherichia coli*: The role of virulence genes. J Microbiol Immunol Infect. 2019 Mar 29;. doi: 10.1016/j.jmii.2019.03.005. [Epub ahead of print] PubMed PMID: 31076319.
103. Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect Immun. 1999 Feb;67(2):546-53. PubMed PMID: 9916057; PubMed Central PMCID: PMC96353.
104. Johnson JR, Orskov I, Orskov F, Goullet P, Picard B, Moseley SL, Roberts PL, Stamm WE. O, K, and H antigens predict virulence factors, carboxylesterase B pattern, antimicrobial resistance, and host compromise among *Escherichia coli* strains causing urosepsis. J Infect Dis. 1994 Jan;169(1):119-26. doi: 10.1093/infdis/169.1.119. PubMed PMID: 7506278.
105. Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. (2000). Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature 406, 64–67.
106. Tartof, S. Y., Solberg, O. D., Manges, A. R., & Riley, L. W. (2005). Analysis of a uropathogenic *Escherichia coli* clonal group by multilocus sequence typing. Journal of clinical microbiology, 43(12), 5860–5864. doi:10.1128/JCM.43.12.5860-5864.2005
107. Gomes, T. A., Elias, W. P., Scaletsky, I. C., Guth, B. E., Rodrigues, J. F., Piazza, R. M., … Martinez, M. B. (2016). Diarrheagenic *Escherichia coli*. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology], 47 Suppl 1(Suppl 1), 3–30. doi:10.1016/j.bjm.2016.10.015
108. Sahl, J. W., Matalka, M. N. & Rasko, D. A. (2012). Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments. Appl Environ Microbiol 78, 4884–4892
109. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Applied and environmental microbiology 2000;66(10):4555-4558
110. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep. 2013 Feb;5(1):58-65. doi: 10.1111/1758-2229.12019. Epub 2012 Dec 24. PubMed PMID: 23757131.

111. 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. J Clin Microbiol. 2015 Jan;53(1):160-6. doi: 10.1128/JCM.02562-14. Epub 2014 Oct 29. PubMed PMID: 25355761; PubMed Central PMCID: PMC4290915.
112. 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. J Clin Microbiol. 2015 Jan;53(1):160-6. doi: 10.1128/JCM.02562-14. Epub 2014 Oct 29. PubMed PMID: 25355761; PubMed Central PMCID: PMC4290915.
113. 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. J Clin Microbiol. 2015 Jan;53(1):160-6. doi: 10.1128/JCM.02562-14. Epub 2014 Oct 29. PubMed PMID: 25355761; PubMed Central PMCID: PMC4290915.
114. Chattaway, M. A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M., … Wain, J. (2014). Evidence of evolving extraintestinal enteroaggregative *Escherichia coli* ST38 clone. Emerging infectious diseases, 20(11), 1935–1937. doi:10.3201/eid2011.131845
115. Johnson JR, Johnston BD, Porter SB, Clabots C, Bender TL, Thuras P, Trott DJ, Cobbold R, Mollinger J, Ferrieri P, Drawz S, Banerjee R. Rapid Emergence, Subsidence, and Molecular Detection of *Escherichia coli* Sequence Type 1193-fimH64, a New Disseminated Multidrug-Resistant Commensal and Extraintestinal Pathogen. J Clin Microbiol. 2019 May;57(5). doi: 10.1128/JCM.01664-18. Print 2019 May. PubMed PMID: 30787145; PubMed Central PMCID: PMC6498021.
116. Majed F. Alghoribi, Tarek M. Gibreel, Garry Farnham, Sameera M. Al Johani, Hanan H. Balkhy, Mathew Upton, Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic *Escherichia coli* clones in Riyadh, Saudi Arabia, Journal of Antimicrobial Chemotherapy, Volume 70, Issue 10, October 2015, Pages 2757–2762, https://doi.org/10.1093/jac/dkv188
117. Veronika L Tchesnokova, Elena Rechkina, Lydia Larson, Kendra Ferrier, Jamie Lee Weaver, David W Schroeder, Rosemary She, Susan M Butler-Wu, Maria E Aguero-Rosenfeld, Danielle Zerr, Ferric C Fang, James Ralston, Kim Riddell, Delia Scholes, Scott Weissman, Kaveri Parker, Brad Spellberg, James R Johnson, Evgeni V Sokurenko, Rapid and Extensive Expansion in the United States of a New Multidrug-resistant *Escherichia coli* Clonal Group, Sequence Type 1193, Clinical Infectious Diseases, Volume 68, Issue 2, 15 January 2019, Pages 334–337.
118. Roy Chowdhury P, McKinnon J, Liu M, Djordjevic SP. Multidrug resistant uropathogenic *Escherichia coli* ST405 with a novel, composite IS*26* transposon in a unique chromosomal location. Front Microbiol 2018;9:3212.

Chapter 3: Methodology

3.1. Collection of strains

In this thesis, whole genome sequencing and analyses was conducted on a collection of 81 *Escherichia coli* samples from blood cultures. These 81 sequences were selected based on the year of isolation, multi-drug resistant phenotype and presence of clinical class 1 integrase gene, which is discussed in details in Chapter 4. These blood cultures were isolated between the year 2013 to 2016 from Concord Repatriate Hospital, Sydney. The cohort was derived from a variety of infections such as biliary tract, soft skin tissue, device related infections, abdominal and urinary tract infections and febrile neutropenia. Each of the blood cultures were isolated from gycerol stocks by growing them on Luria Broth (LB) agar plates at the Concord Repatriate General Hospital, Sydney.

The *E. coli* strains sequenced and analysed in this thesis was a part of a pilot study that provided the basis to examine antimicrobial resistance and virulence in BSI isolated from a hospital in Australia. It is important to note that this collection of isolates revealed highly diverse *E. coli* clones that varied in their ARG and virulence patterns, further discussed in Chapter 4. The sample size for genomic characterisation is small comprising of only 81 isolates and may not reveal the true diversity of Australian population. However, the entire cohort of 406 collected along with the meta-data analysis is provided in Chapter 4 for comparison and co-relations with other published data in Australia and world-wide.

3.2 Genomic epidemiology:

Genomic surveillance is a powerful tool to track and characterise pathogens and to understand the dissemination of antimicrobial resistance and *E. coli* clones in human, animal and environmental settings. Genomic epidemiology is now widely used for identifying infection outbreaks and provides the ultimate resolution between two or more closely related pathogens. It is also useful to rapidly identify the source of infection and provides links of outbreaks to environment and other sources. The rapid advances in whole genome sequencing (WGS) has made it a cost effective method replacing traditional phenotyping methodologies in microbiology laboratories. Within the last decade whole genome sequencing has been increasing used as a monitoring system to detect the emergence and dispersal of clinically important antimicrobial resistant bacterial populations [1] [2].

Although clinical genomics provides a higher confidence of confirmation of outbreaks, characterising MDR clones, and phylogenetic relationships between clinically important strains, it still has certain limitations that are important to be addressed. Open sourced softwares are not always very robust and it is essential to use peer reviewed softwares for genomic data analysis. In addition, there is no standard workflow or protocols for conducting genomic epidemiology and the tools and softwares used for this may differ depending on the research questions with a particular project [3].

With microbiology labs, updated genome sequence equipment and upgrades for bioinformatics tools are not always accessible. Moreover, there is lack of highly trained staff that can successfully analyse genomic infromation, this is, co-relate genomic information with the phenotypic data and translated back to infection control via effective antimicrobial stewardship. Other obstacles also include maintaince and implementation costs for ensuring high standards and high quality for conducting genomic epidemiology [3].

However, in the future with WGS and the availabilty of a number sequence databases such as GenBank and data sharing across the globe, epidemiological analysis would provide the foundation for antimicrobial stewardship and successful interventions for clonal outbreaks.

3.3 Research Methodology:

Each of the result chapters include the full description of the methods and materials. This section will comprise of an overview of the methodology of this thesis.

3.3.1 DNA isolation:

DNA isolation was conducted using the Bioline ISOLATE II Genomic DNA Kit. The isolated DNA was used to prepare whole genome libraries for short read sequencing. DNA isolation for long read sequencing was conducted using phenol:chloroform extraction technique in order to provide high quality, pure and high concentration of DNA. This method was gentler than the Bioline kit as there was less chance of shearing of DNA.

Short read whole genome sequencing was conducted on all 81 *E. coli* isolates using the Illumina Hi-Seq 2500 system, that uses cell-flow technology for rapid high-throughput sequencing (https://sapac.illumina.com/systems/sequencing-platforms/hiseq-2500). Long read sequencing was performed on a subset of *E. coli* isolates, discussed in the result chapters. Long

read sequencing platforms that were used in this thesis were; 1) PacBio RS II sequencer with Single Molecule Real Time (SMRT) sequencing (https://www.pacb.com/); 2) the MinION sequencer using Nanopore technologies (https://nanoporetech.com).

3.3.2 Sequence assembly:

Short read assemblies were generated using A5-mi seq platform that are analysed in the result chapters. The short read assemblies provided the genetic content to characterise antimicrobial resistance regions within the cohort but the disadvantage of this was that the complete assembly of these regions was not possible with short reads.

The long reads were passed through Unicycler software that generates hybrid assemblies with the nanopore or pac-bio raw reads and the Illumina raw reads. Unicycler generates a short read assembly and then utilises the long reads to bridge the gaps and this typically generates closed plasmids and chromosomes. This is followed by the polishing step to ensure that the base calling of the final contigs is accurate [4]. The visualisation of these hybrid assemblies was conducted using Bandage software [5].

3.3.3 Strain identification and phylogeny:

Multi-locus sequence typing (MLST):

MLST of *E. coli* is conducted using a number of schemes, we used the Achtman MST scheme for our cohort. This scheme was based on 7 housekeeping genes that categories E. coli into a unique sequence type based on different allelic combinations. It is popularly used in population genomics for identifying bacterial genotypes, that includes clones and clonal complexes that can be co-related to biological properties seen within them, such as phenotype, pathogenicity and bacterial evolution [6] [7]. We implemented pubMLST to identify sequence types within our cohort.

3.3.3.1 Serotyping and phylogrouping:

Serotyping is a method for classification of *E. coli* developed in the 1940s traditionally based on variation of cell surface antigens that includes H flagellar antigen and O lipopolysacchride antigen. In contrast to traditional serotyping, molecular serotyping that is a genetic assay based on targetting O-group specific genes present in *E. coli* O-antigen and H antigen genes that differ depending on different flagella. One of the most recently popular molecular serotyping method is in-silico serotyping that uses whole genome sequences. In this study we used the

Web tools available from Center for Genomic Epidemiology (CGE) (www.genomicepidemiology.org) for detecting the serotype. Phylogrouping was detected similarly based on the three housekeeping genes that categories *E. coli* into five different groups. We conducted this using the assembled sequences generated by A5 pipeline [8].

3.3.3.2 Marker genes based phylogeny:

Phylosift is a pipeline was intially utilised for metagenomic analysis. This pipeline that uses 37 conserved prokaryotic marker gene families to extract, translate and align sequences and this provides an insight into the phylogenetic relationship between strains. This method is useful as it does not depend on a reference genome for alignment [9].

3.3.3.3 SNP based phylogeny:

Parsnp is a powerful tool for determining phylogenetic relationships based on single nucelotide polymorphisms (SNPs). It uses a reference sequence to generate high resolution SNP phylogeny based on the core-genome. It is an excellent method to determine genetic relatedness between strains of the same MLST or clonal complex. We used this method along with recombination filters to align the strains against the reference genome. Parsnp was utilized in the result chapter 5 and 6 [22].

3.3.3.4 Mr Bayes based phylogeny:

This method was used for determining the molecular evolution of closely related strains using General Time Reversal (GTR) model to generate a SNP phylogeny using 300 trillion iterations of bootstrapping [11].

3.3.4 Gene identification:

Easy and rapid detection of genes of interest can be conducted using either assembled genomes or short reads.

3.3.4.1 Based on assembled genomes:

National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) is implemented online and can also be used on the command line [12]. BLASTn is a sequence alignment tool that align nucelotide sequences against reference sequences. We used this tool on the command line to screen our cohort for genes of interest. GenBank nucleotide database is an annotated collection of publicly available DNA sequences that can be accessed

via BLAST. Other online databases that were used included Phage Search tool (PHAST), to identify any phage associated sequences [13]; IS finder to identify putative IS elements (https://www-is.biotoul.fr/) and Open Reading Frame Finder, used to identify open reading frames putatively associated with the coding regions (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).

3.3.4.2 Based on short reads:

Short reads were used for identifying genes as this prevented the assembly bias. ARIBA (Antimicrobial Resistance Identification by Assembly) was used with gene databases, that included ResFinder, VFDB and PlasmidFinder [19-20] [16].

3.3.4.3 Genome annotations:

Characterisation of plasmids and antimicrobial resistance regions has been conducted in Chapter 5 and 6. For this we used Rapid Annotation using Subsystems Technology toolkit (RAST) via the Pathosystems Resource Integration Centre's (PATRIC) online Bacterial Bioinformatics Resource Centre [17].

3.3.5 Comparative genomics and data visualisation:

3.3.5.1 Snapgene:

Snapgene v3.3.4 is package for gene annotation, visualisation and management of genomes (https://www.snapgene.com/). In our result chapters we have used this tool for generating high quality figures with annotations.

3.3.5.2 Mauve:

Progressive Mauve is a multiple genome alignment package that is excellent for comparative genomic studies to analyse multiple closely related genomes in one time. This tool identifies conserved segments during alignment that are typically free from genome rearragements or recombinations [24].

3.3.5.3 Easyfig:

Easyfig is a Python based tool for conducting comparision figures of multiple genomes using a user friendly graphic interface [14]. This tool was useful for identifying differences in plasmids and antimicrobial resistance regions in the result chapters.

3.3.5.4 BLAST ring generator (BRIG):

This is an excellent visualisation tool that was used for detecting similarities and differences of a number of genomes against a central reference sequence. BRIG performs BLAST comparisions and generates a ring figure that includes the read coverage, contig boundries and customised annotations [18].

3.3.5.5 Data visualisation:

FastTree2 that works under the generalised time-reversible (GTR) model, to generate maximum likelihood trees from alignments [15]. We used this tool for inference of our phylogenetic trees generated in all of the result chapters. R package ggtree is another tool we used to conduct visualisation and manipulation of phylogenetic trees and was also used to generate heatmaps, shown in the result chapters [21].

3.4 References:

1. Luheshi, L.M., Raza, S. & Peacock, S.J. Moving pathogen genomics out of the lab and into the clinic: what will it take?. *Genome Med* **7,** 132 (2015) doi:10.1186/s13073-015-0254-z
2. Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM et al. Genomic Microbial Epidemiology Is Needed to Comprehend the Global Problem of Antibiotic Resistance and to Improve Pathogen Diagnosis. Front Microbiol 2016;7:843.
3. Tang, K. L., et al. (2017). "Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis." Lancet Planet Health 1(8): e316-e327.
4. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 2017;13(6):e1005595.
5. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 2015;31(20):3350-3352.
6. Wirth T, Falush D, Lan R, Colles F, Mensa P et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Molecular Microbiology 2006;60(5):1136- 1151.
7. Robins-Browne, Roy M et al. "Are *Escherichia coli* Pathotypes Still Relevant in the Era of Whole-Genome Sequencing?." *Frontiers in cellular and infection microbiology* vol. 6 141. 18 Nov. 2016, doi:10.3389/fcimb.2016.00141
8. Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. J Clin Microbiol, 53(8), 2410-2426. doi:10.1128/jcm.00008-15
9. Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2014;2:e243.
10. Treangen TJ, Ondov BD, Koren S, Phillippy AM: Rapid Core-Genome Alignment and Visualization for Thousands of Microbial Genomes. bioRxiv (2014). doi: http://dx.doi.org/10.1101/007351
11. Huelsenbeck J.P., Ronquist F. (2005) Bayesian Analysis of Molecular Evolution Using MrBayes. In: Statistical Methods in Molecular Evolution. Statistics for Biology and Health. Springer, New York, NY
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403-410.
13. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011). PHAST: a fast phage search tool. Nucleic Acids Res 39, W347-352.
14. Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009-1010.
15. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 2010;5(3):e9490.
16. Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom 2017;3(10):e000131.
17. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 2015;5:8365.
18. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics, journal article 2011;12(1):402.

19. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 2014;58(7):3895-3903.
20. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012;67(11):2640-2644.
21. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y et al. ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 2017;8(1):28-36.
22. Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15, 524.
23. Darling AE, Mau B, Perna NT: progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one 2010, 5(6):e11147.

Chapter 4: Genomic profiling of *Escherichia coli* isolates from bacteraemia patients: A 3-year cohort study of isolates collected at a Sydney teaching hospital

Priyanka Hastak[1,2], Max L Cummins[1], Thomas Gottlieb[3,4], Elaine Cheong[3], John Merlino[3,4], Garry S. A. Myers[1], Steven P. Djordjevic[1,2] and Piklu Roy Chowdhury[1,2*]

[1]The ithree institute, University of Technology Sydney, City Campus, Ultimo, NSW 2007, Australia.

[2] Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia

[3]Department of Microbiology and Infectious Diseases, Concord Hospital, NSW Health Pathology, Hospital Road, Concord 2139, NSW, Australia.

[4] Faculty of Medicine, University of Sydney, NSW Australia.

4.1.2 Data Accessibility:

Genome sequences of isolates included in this manuscript have been made accessible via GenBank under project ID PRJNA480723. A list of individual accession numbers for the assembled genomes is provided in Supplementary File S2_Chapter4. The genome sequence of the novel *E. coli* ST8196 was registered in Enterobase and is searchable using the ST number.

4.2 Abstract

Objectives: This study sought to assess the genetic variability of *Escherichia coli* isolated from blood-stream infections (BSI) presenting at Concord Hospital, Sydney during 2013-2016.

Methods: A cohort of 406 samples were collected and resuscitated by PH from freezer collection at Concord Hospital Sydney and transported pure cultures on LB agar plates to iThree institute, UTS. Whole genome sequencing was used to characterise 81 *E. coli* isolates sourced from community onset (CO) and hospital onset (HO) blood-stream infections.

Results: The cohort comprised 64 CO and 17 HO isolates, including 35 multidrug resistant (MDR) isolates exhibiting phenotypic resistance to three or more antibiotic classes. Phylogenetic analysis identified two major ancestral clades. One was genetically diverse with 25 isolates distributed in 16 different sequence types (STs) representing phylogroups A, B1, B2, C and F, while the other comprised phylogroup B2 isolates in subclades representing ST131, ST73 and ST95 lineages. Forty-seven isolates contained a class 1 integron, of which 14 carried $bla_{\text{CTX-M}}$. Isolates with a class 1 integron carried more antibiotic resistance genes than isolates without an integron and in most instances, resistance genes localised within complex resistance loci (CRL). Resistance to fluoroquinolones could be attributed to point mutations in *parC* and *gyrB* genes and two isolates carried a plasmid-associated *qnrB4* gene. Co-resistance to fluoroquinolone and broad-spectrum beta-lactam antibiotics was associated with ST131 (HO and CO), ST38 (HO), ST393 (CO), ST2003 (CO) and ST8196 (CO and HO), a novel ST identified in this study. Notably, 10/81 (12.3%) isolates with ST95 (5 isolates), ST131 (2 isolates), ST69, ST88 and ST549 (1 isolate each) carried an IncFII-IncFIB plasmid replicon and virulence genes consistent with the carriage of ColV-like plasmids.

Conclusions: Our data indicated that *E. coli* causing BSI were genetically diverse and belonged to a variety of ST's. They carried a number of clinically important antimicrobial resistance and virulence associated genes.

4.3 Introduction:

Escherichia coli is a major cause of bloodstream infections (BSI), dominating the list of Gram-negative pathogens associated with bacteraemia and sepsis globally [1]. In 2017, treatment costs for patients with sepsis in Australian intensive care units was estimated to be at least $846 million [2]. In that year 55.2% of Gram-negative BSIs were caused by *E. coli,* of which 83.6% were from community and 16.4% from hospitalised patients [3]. Nationwide, the incidence of multidrug resistance (MDR) associated with *E. coli* BSI increased from 11.8% in 2013 [4] to 21.9% in 2017 [3]. This was also accompanied by an increase in the frequency of isolation of MDR-*E. coli* displaying resistance to fluoroquinolones and extended spectrum beta-lactams [3].

Extra-intestinal pathogenic *E. coli* (ExPEC) are usually associated with bloodstream infections and sepsis. ExPEC causing human infections are often characterised by the presence of P and S fimbriae that facilitate human cell attachment, toxins and cytonecrotic factors that enable invasion and subsequent damage, iron acquisition systems, serum resistance proteins [5, 6], capsular polysaccharides and lipopolysaccharides (LPS) that support survival in the host. *E. coli*-BSI predominantly have a community onset, with urinary tract infections (UTI) being the most frequently reported principal clinical manifestation [3, 7]. Uropathogenic *E. coli* (UPEC) are a subgroup of ExPEC that are adept at colonising the urethra and ascending into the urinary bladder causing cystitis. A three-month sentinel surveillance of *E. coli* bacteraemia cases in the UK in 2012/2013, which included 35 hospitals, linked 51.2% of the cases to UTI [8]. The first-line drugs used to treat UTI in Australia include trimethoprim, augmentin and cephalexin. Genes conferring resistance to trimethoprim and cephalexin are often present as gene cassettes on class 1 integrons, a globally disseminated mobile genetic element best known for its role in the rapid evolution of multi-drug resistant pathogens [9-11]. In Gram-negative genomes, class 1 integrons are frequently detected in regions where resistance genes cluster together, forming resistance loci. Hence they serve as a reliable proxy for identification of multiple drug resistant genomes within cohorts of bacteria collected from clinical settings [12], food animals [13-15] and the environment [16]. Increasing resistance to aminoglycosides, extended spectrum beta-lactams, fluoroquinolones and trimethoprim is a major threat to the treatment of *E. coli* infections including BSIs [17-21]. Plasmids often play a major role in driving the evolution of pathogens resistant to these antibiotics through the rapid dispersal of virulence and antibiotic resistance genes [22, 23] within Enterobacterales. Examples of plasmid-mediated rapid pathogen evolution include the establishment of Enterotoxigenic *E. coli* (ETEC) [24, 25] and

Enteroinvasive *E. coli* (EIEC) [26] pathotypes, which cause watery diarrhoea and bacillary (often bloody) dysentery respectively. The epidemic of HUS in 2011 caused by MDR O104:H4 highlighted the power of pathogen genomics in rapidly identifying the cause of treatment failure and evaluating alternative treatment options [27].

Globally, the most predominant ExPEC-sequence types (ST) associated with UTI include ST69, ST73, ST95 and ST131 [28, 29]. Specifically, the ST131 lineage is recognised as a significant etiological agent of UTI, pyelonephritis and urosepsis [30, 31]. The wide distribution of ST131, together with high drug resistance and virulence has made it the most well studied ExPEC lineage [32, 33]. The emergence of MDR-BSI is often associated with isolates that have acquired genes encoding extended spectrum beta-lactamases (ESBLs), particularly *bla*$_{\text{CTX-M.}}$ There are over 100 allelic variants of *bla*$_{\text{CTX-M}}$ [34-36] and different alleles predominate in *E. coli* depending on geographic location [37] and other factors. BSI in Argentina, South Africa, Turkey, Belgium, China, Taiwan and Australia are typically caused by *E. coli* which carry *bla*$_{\text{CTX-M-2}}$, *bla*$_{\text{CTX-M-3}}$ and *bla*$_{\text{CTX-M-15}}$ [38-40].

Surveillance of antibiotic resistant pathogens has provided important insights into the distribution and prevalence of antibiotic resistance in Australia [41], and has highlighted rising rates of ESBL-producing and fluoroquinolone-resistant MDR-*E. coli* causing BSIs [3, 7, 42]. Whole genome sequencing of pathogens causing BSI facilitates identification of AMR gene carriage and virulence-associated genes (VAGs) and links their carriage with ST, phylogroup and e-serotype. It also provides insights into the genetic context of AMR genes and VAGs and sheds light on associations of AMR genes with mobile genetic elements. In addition to these attributes, WGS enables effective microbial source tracking.

In this report, we present an analysis of *E. coli* whole genome sequences collected as part of the AGAR Gram-negative Sepsis Outcome Program (GNSOP) from patients with bacteraemia (BSI) attending Concord Hospital, Sydney between January 2013 and March 2016. We investigated the genetic diversity of *E. coli* that cause BSI, describe the genetic context of resistance gene carried by these isolates and identify genetic elements that disseminate AMR genes and VAGs.

4.4 Methods:

4.4.1 *Escherichia coli* collection and growth conditions:

Between January 2013 and March 2016, 406 *E. coli* isolates were collected by the Microbiology Department at Concord Repatriation Hospital in Sydney for the AGAR-GNSOP survey. We selected 81 representative isolates over the spectrum of phenotypic antibiotic resistance profiles for whole genome sequence analysis. Clinical metadata including collection year, likely source of infection and phenotypic resistance profile is presented in **S1_Chapter4**. A 48-hour post-admission cut-off for sample collection was used to distinguish community onset (CO) and hospital onset infections (HO). Isolates were sub-cultured and grown at 37^{o}C for 18 hours on Lysogeny Broth (LB) and the pellets were used for purification of DNA.

4.4.2 Antimicrobial resistance phenotyping:

Antibiotic susceptibility profiles were generated using the Advanced Expert System (AES) software designed for VITEK2 (BioMerieux). Breakpoints were interpreted using EUCAST (European Committee on Antimicrobial Susceptibility Testing) [43] at the Microbiology Laboratory, Concord Hospital. The panel of antibiotics tested comprised ampicillin (AMP), amoxicillin (AMC), ciprofloxacin (CIP), cefazolin (CFZ), cefotaxime (CTX), ceftriaxone (CRO), ceftazidime (CAZ), cefoxitin (FOX), trimethoprim (TRI), gentamicin (GEN), tobramycin (TOB), norfloxacin (NOR), ticarcillin/clavulanic acid (TCC), piperacillin/tazobactam(TZP) trimethoprim-sulfamethoxazole (SXT) and amikacin (AMK). Synergy tests were used to phenotypically identify isolates producing ESBL enzymes (**S1_Chapter4**). For the purposes of this study, multi-drug resistance (MDR) was defined as resistance to ≥3 classes of antibiotics [44].

4.4.3 Genomic DNA extraction and sequencing library preparation:

Genomic DNA was extracted from 2mL of overnight cultures using ISOLATE II Genomic DNA Kit (Bioline, Australia), following the manufacturer's protocol, and quantified using the Qubit fluorimeter and dsDNA HS Assay Kit (Thermo Fisher Scientific, Australia). Whole genome sequencing libraries were prepared from 2ng of gDNA using the Illumina Nextera DNA kit following previously published protocols [45].

4.4.4 Whole genome sequencing and assembly:

Whole genome sequencing was performed using the Illumina HiSeq 2500 v4 using rapid PE150 mode (Illumina, San Diego, CA, USA), and assembled using the A5-miSeq assembly

pipeline [46]. The number of scaffolds per genome varied from 30 to 200, with median sequence coverage ranging between 30× and 55× (File S2_Chapter4). Preliminary genome annotations were performed on RASTtk (http://rast.nmpdr.org/) [47] and regions of interest were manually curated using BLASTn and iterative BLASTp searches. The assembled genome data are in GenBank under project ID PRJNA480723. The BioSample accession numbers are listed in File S2_Chapter4.

4.4.5 Gene identification and MLST analysis:

The Antimicrobial Resistance Identification By Assembly (ARIBA) pipeline [48] was used to identify the class 1 integrase gene *intI1*, antimicrobial resistance genes (ARGs), virulence-associated genes (VAGs), specific insertion elements, O and H-antigen genes, plasmid replicon types and Clermont Phylogroups. Raw Illumina reads were mapped to pCER4 (accession no. KU578032) and pSDJ2009-52F (accession no. MH195200.1) to identify genetic markers present in IncFII-IncFIB-ColV hybrid plasmids using published protocols [13]. Multi-locus sequence type (MLST) analysis of plasmids was performed using the Achtman *E. coli* MLST scheme on PubMLST database (http://pubmlst.org/).

4.4.6 Phylogenetic Analysis:

Reference gene-based maximum-likelihood phylogeny analysis (using 37 prokaryotic marker genes) was performed using Phylosift [49]. FastTree version 2.1.8.c was used to construct a phylogenetic tree (http://www.microbesonline.org/fasttree/FastTree-2.1.8.c) and visualised using FigTree v1.4.2 [45]. Reference genomes used in the analysis were *E. coli* K12 (U00096.2) and *E. coli* EC958 (NZ_HG941718.1); *Klebsiella pneumoniae* subsp. *pneumoniae* MGH78578 (CP000647.1) was used as an outlier. To increase the resolution of the branches, the outer layer was removed from the final tree presented in this manuscript (**Figures 1 and 2**). The tree figures were edited using iTol online tree management software (https://itol.embl.de/) and virulence heat maps were created using ggplot2 in *R*-package [50].

4.5 Results:

4.5.1 Isolate metadata:

Clinical metadata corresponding to the 81 de-identified isolates (**S1_Chapter4**) indicated that 37 isolates (45.5%) were associated with UTIs and most of these (64/81; 79%) were recovered from community onset infections. The cohort included 35 MDR isolates of which 13 exhibited

co-resistance to fluoroquinolones and extended spectrum beta-lactams. Assembly statistics (File S2_Chapter4) indicated ≥24 x coverage, except EC107_ST69 (19 x coverage).

4.5.2 Phylogeny and Sequence Type distribution:

Thirty different *E. coli* sequence types (ST) (**S1_Chapter4**) were identified within this BSI cohort. Phylogenetic analysis using Phylosift revealed two major ancestral clusters (**Figure 1**). The upper major clade contained 25 isolates distributed in 16 STs (**Figure 1**). Fifty-six isolates that clustered in the lower major clade comprised 15 STs, including globally recognised *E. coli* clonal groups ST73 (17isolates), ST131 (15 isolates), ST95 (8 isolates) and ST127 (4 isolates). Isolates in the upper major clade of the phylogenetic tree were represented by Clermont phylogroups B2, D, F, A and B1 while all isolates in the lower major clade of the phylogenetic tree, irrespective of ST type and O and H surface antigens typed as phylogroup B2. The relative distribution of the clinically significant STs within CO infections were ST73 (16/17), ST131 (10/15), ST95 (7/8) and ST69 (2/5). Resistance genes, class 1 integron associated genes and plasmid replicon types in the entire cohort is presented in **Table1_Chapter4**.

E. coli sequence types ST8196 (2 isolates) and ST8197 (one isolate) are novel and were registered with Enterobase (https://enterobase.warwick.ac.uk/species/index/ecoli). ST8196 is closely related to the ST131-O25:H4 sub-clade (**Figure 1**) and is a new member of the ST131 clonal complex. The ST8196 isolates, EC233 and EC234, were MDR and were isolated within a 24 hour period from separate patients. EC233 was from a patient with community onset UTI, while EC234 was from a patient with a hospital-acquired infection. Both isolates displayed phenotypic resistance to fluoroquinolones and expressed extended spectrum β-lactamases. The ST8197 isolate EC128 is closely related to members of the ST144 subclade (**Figure 1**), a sequence type which shares a close common ancestor with ST95. Isolate EC128_ST8197 was resistant to ampicillin and trimethoprim-sulfamethoxazole only.

Most (21; 60%) of the 35 MDR isolates clustered in the ST131 (13 isolates) and ST73 (8 isolates) subclades (**Figure 1**). Genes responsible for ESBL production and fluoroquinolone-resistance phenotypes within the MDR subset, were not restricted to isolates in the ST131 sub-clade but were observed in ST38 (HO isolate EC36), ST393 (CO isolate EC66), ST2003 (CO isolate EC137) and ST8196 (EC233 from CO and EC234 from HO).

4.5.3 Genetic context of Class 1 integrons and resistance genes:

Of the 81 isolates, 47 (58%) carried a class 1 integron (Table 1). Aminoglycoside adenyltransferase genes *aadA1*, *aadA2* (streptomycin resistance) and *aadA5* (spectinomycin resistance), dihydrofolate reductase gene alleles *dfrA1, dfrA7, dfrA14, dfrA14* and *dfrA17*, encoding resistance to trimethoprim, *cmlA1* (encodes chloramphenicol resistance) and $bla_{\text{OXA-1}}$ (encodes host specific low level resistance to ESBL antibiotics) were a feature of cassette arrays associated with class 1 integrons in this collection (Table 1). Fifteen isolates harboured the *aadA5-dfrA17* (**Figure 2A**) cassette array, while others comprised different combinations of the gene cassettes listed above and in Table 1. Genomic scaffolds containing class 1 integrons often carried other clustered resistance genes (**Figure 2**) interspersed with different insertion elements. Integrons bearing the *aadA5-dfrA17* cassette array were almost always linked to a macrolide resistance module encoded by *mphA* and *mphR* genes beyond the 3′-CS, forming a complex resistance locus (CRL) (**Figure 2A**). IS*26* sequences flanked the majority of scaffolds carrying a class 1 integron (**Figure 2**). Notably most isolates (64/81; 79%), irrespective of resistance gene carriage, carried multiple copies of IS*26* (**S1_Chapter4**).

The genes encoding resistance to ESBL were restricted to three allelic variants of $bla_{\text{CTX-M}}$: $bla_{\text{CTX-M-14}}$, $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$. In most instances $bla_{\text{CTX-M-15}}$ genes were associated with IS*Ecp1* (**S1_Chapter4**). However, in EC137_ST2003 and EC398_ST4702, $bla_{\text{CTX-M-14}}$ was associated with IS*Ecp1b.* Isolate EC345_ST10 had a class 1 integron with a $bla_{\text{OXA-1}}$ gene cassette (**Figure 2I)**, but with no detectable enzymatic activity. Fluoroquinolone resistance in most isolates was attributed to mutations in *gyrB* and *parC* except in isolates EC46_ST131, EC44_ST131 and EC137_ST2003, which additionally had acquired *aac(6')-1b-cr*, a gene known to confer low level resistance to fluoroquinolones and aminoglycosides. Fluoroquinolone resistance encoded by *qnrB4* was identified in isolates EC233_ST8196 and EC234_ST8196.

4.5.4 *E. coli* phylogroups and virulence gene profiling:

The 81 genomes were screened for a panel of 58 VAGs found in *E. coli* (**Figure 3, Table S3_Chapter4**). The relative abundance of VAGs was highest in ST95 and ST73 isolates, followed by ST131 (Figure 3, File S4). As expected, isolates belonging to Clermont phylogroups A, B1 and F carried fewer VAGs than the isolates in Phylogroups B2 and D (**Table S4_Chapter4**). The *fimH*, *yeeT*, *fyuA*, *irp2, sitA* and *iss* genes were identified in over 80 percent of the isolates in our cohort (**Table S4_Chapter4**). Only four isolates EC104, EC2,

EC230 and EC19 (all phylogroup B2) carried *ibeA*. Of the two iron acquisition genes, *iucD* and *irp2*, *iucD* was present in 52 isolates (8 phylogroup D and 1 each of phylogroup A and D and the rest B2), and 72 tested positive for *irp2* (all 68 phylogroup B2 and two each of A and F). A previous single hospital study of *E. coli* from BSIs highlighted an abundance of *fimH*, *papC*, *iha* and *ibeA* [51].

The distribution of the six ExPEC specific protectins, namely *kpsMT II, kpsMT III, traT*, *ompT, iss* and *cavC* and two toxins, *hly* and *cnf1,* was also determined. Forty isolates tested positive for *kpsMT II,* while only isolate EC323 tested positive for *kpsMT III.* 14 isolates were positive for *ompT*, 58 for *traT*, 67 for *iss* and 11 for *cavC* (**Table S3_Chapter4**). None of the isolates tested positive for *hlyA*, but 22 (all phylogroup B2) were positive for *cnf1*. The distribution of genes encoding known outer membrane iron receptors like *fyuA, ireA, iroN* and *iutA* was random amongst the *E. coli* phylogroups (**Table S4_Chapter4**).

4.4.5 Plasmid replicon diversity within the cohort:

Plasmid typing identified IncFII-IncFIB replicons in most isolates (64/81) including 41 CO-BSI and 12 HO-BSI isolates (Table 1). In 11 isolates from CO infections and 3 from HO infections FIA replicons were co-resident with IncFII-IncFIB replicons. We also identified a col-like plasmid replicon in 41 of the 64 isolates which carried an IncFII-IncFIB replicon, raising the possibility of the presence of IncFII-IncFIB-ColV virulence plasmids similar to pCERC4 (File S5A) or pSDJ2009-52F (File S5B) plasmids that were reported recently in Sydney [52, 53]. Raw Illumina reads of all 81 genomes were mapped against both plasmids and identified 10 isolates which carried most virulence and colicin resistance genes characteristic of ColV-virulence plasmids (Figures S5A and S5B). Of these, five typed as ST95, 2 as ST131, and 1 each as ST69, ST88 and ST549. However, not all of the 10 isolates carried both ColV and ColIa colicin operons seen in pCERC4 and pSDJ-52F, indicating subtle variations in plasmid structure. Further studies are required to examine the distribution of ColV-like VAGs in these strains.

We also identified an IncI1 replicon in six isolates (4 associated with HO and 2 with CO infections), four of which had an IS*Ecp1* associated with $bla_{\text{CTX-M-15}}$ (Table 1). An IncX4 replicon was identified in four isolates, three of which were ST131. Finally, an IncQ replicon was identified in three MDR-CO isolates belonging to ST38, ST349 and ST88.

4.5 Discussion:

This study and others show that bloodstream infections are caused by genetically diverse *E. coli* [54], indicating that pathogen virulence profiles and perhaps host immune status are likely to be important variables in disease causation. *E. coli* ST131, often isolated from patients with BSI, is widely disseminated in Europe, Asia, North America, Africa and Australia [33, 55] and has played a critical role in the global spread of $bla_{\text{CTX-M-15}}$ [56]. Our data shows that $bla_{\text{CTX-M-15}}$ in Australian BSIs is not restricted to ST131 isolates but extends to *E. coli* ST8196, ST393 and ST38. Chromosomal $bla_{\text{CTX-M-15}}$ in ST131 is usually associated with IS*EcpI* elements, however in EC44_ST131, EC46_ST131 and EC66_ST393 $bla_{\text{CTX-M-15}}$ is not linked to IS*EcpI* suggesting a different genetic context, including an association with other plasmids.

Although our data is based on a limited number of isolates acquired from a single hospital in Sydney, it highlights similar population dynamics in the context of resistance and virulence genes found in invasive isolates reported on a cohort of more than 1500 *E. coli* collected over 11years in the UK by Kallonen et al [57]. The number of virulence genes carried by ST131 isolates in our cohort was also less than in the cohort of *E. coli* ST73 isolates (**Figure 3 and Table S3_Chapter4**), suggesting that there is segregation between excessively virulent and extremely drug resistant clonal lineages [57]. Increased mortality is typically associated with multi-drug resistance and treatment inadequacy, it is therefore important to detect these highly resistant clonal lineages to ensure effective clinical treatment. ST73 was reported as a highly virulent clonal lineage in this collection as most of these VAGs are known to be associated the invasiveness, colonization and host adherence, they influence the overall fitness of the clonal lineage. However, acquisition of virulence plasmids such as ColV that also carry CRL may alter this scenario, necessitating constant surveillance to identify emerging resistance threats [52]. While lineages of ST73 with a MDR genotype have been reported in Australia, drug sensitive and lineages devoid of plasmids are also well described [58].

IncFII-IncFIB plasmid replicons predominated in genetically diverse isolates acquired from both CO and HO BSI. Virulence gene profiling studies suggested that a subset of *E. coli* isolates with ST95, ST131 and others carry IncFII-IncFIB-ColV plasmids. Plasmids which have the ColV operon associated with IncF transfer regions often harbour a combination of virulence genes linked to serious human and animal disease [52] [59-63]. Based on an animal model of UPEC disease, the acquisition of ColV-like plasmid-associated virulence genes is significant in disease aetiology [64, 65]. In addition, the acquisition of a ColV-like plasmid by *Salmonella*

enterica serovar Kentucky has been purported to be a key genetic event that transformed its ability to cause serious disease in poultry by conferring new colonization and fitness capabilities [61]. ColV plasmids play significant roles in invasive avian diseases such as colibacillosis caused by avian pathogenic *E. coli* (APEC) in commercial poultry and are considered a defining trait of this ExPEC pathotype [13, 59]. ColV plasmids carry an extensive suite of VAGSs necessary for colonisation and disease formation in diverse hosts including humans. Gene products that mediate increased serum survival (*iss*), membrane vesicle formation protein (encoded by *hlyF* gene), outer membrane protease activity (encoded by *ompT*), aerobactin biosynthesis (*iuc* operon), temperature sensitive hemagglutination (encoded by *tsh*) and various (*sit, iroN* and *eit*) iron acquisition systems [5, 53] are significant in this regard. Mapping of Illumina reads against two hybrid IncFII-IncFIB-ColV plasmids confirmed the presence of genes in the ColV operon in EC11_ST95, EC152_ST95C, EC323_ST549H, EC337_ST69H and the ColIa operon, often found associated with the ColV operon in ColV virulence plasmids, in isolates EC104_ST95C, EC177_ST95H, EC68_ST95C, EC12_ST131H, EC120_ST131C, EC337_ST69H, EC236_ST540C and EC323_ST549H. Collectively our data indicates that there is variability in the IncF plasmids circulating in our BSI cohort [52]. Further studies using long read sequencing methodologies will address the diversity of plasmids carrying ColV-like virulence gene cargo and examine their antibiotic resistance gene cargo.

Forty-seven isolates in the bacteraemia cohort (58%) carry a class 1 integron (Table 1). As expected, most of the *intI1* positive isolates carried a larger number of antibiotic resistance genes compared to isolates that did not have a class 1 integron. Almost all *intI1* carrying *E. coli* carried a minimum gene set encoding resistance to streptomycin (*aadA* gene cassettes) and/or trimethoprim (*dfrA* gene cassettes), sulphonamide drugs (*sul1*, *sul2* or *sul3* genes) and ampicillin ($bla_{\text{TEM-1}}$). Most isolates (34/46; 73%) displayed phenotypic resistances to trimethoprim and sulphonamide drugs, which are used as first line treatments for UTI. The class 1 integron is arguably one the most successful genetic elements responsible for the rapid dispersal of multiple antibiotic resistance genes as it has captured a wide array of resistance gene cassettes [66], colonised diverse animal and plant hosts and has infiltrated every continent [67]. Its structure continues to change, often mediated by insertion elements such as IS*26* [68-71] that can mobilise diverse antibiotic resistance genes. The class 1 integron is often found on diverse plasmid backbones including conjugative broad host range plasmids, and it is widely considered to be a reliable proxy for multiple drug resistance [16, 72].

The MDR isolates in this cohort exhibited little diversity in the carriage of cassette-associated genes. Most class 1 integrons were variants of In2 or In4, some of which were associated with Tn*21*/Tn*1721* hybrid transposon backbones. Our data identified IS*26* adjacent to resistance genes in most of the CRLs (Figure 3). We also provide evidence for multiple IS*26* elements per *E. coli* genome (Table 1), reinforcing the role played by IS*26* in driving the evolution of CRL in the Enterobacterales including those causing BSIs. This is a critically important observation because: i) IS*26* can mobilise a wide range of antibiotic resistance genes; ii) IS*26* is a preferred site for integration of translocatable units (TU's) flanked by IS*26*; and iii) CRLs flanked by IS*26* have the capacity to move laterally as independent TU's [73]. These observations emphasise the necessity for developing and implementing comprehensive surveillance systems for resistance gene epidemiology. While the acquisition of real-time surveillance data on pathogen genomes in clinical arenas is critical for determining the choice of antibiotics for patient treatment, improvements are needed to better frame infection control strategies to identify hotspots for the evolution of MDR pathogens.

One of the most significant findings from this study is the identification of ST8196 within the ST131 clonal complex. ST8196 isolates shared similar genotypic, phenotypic and plasmid profiles with ST131 isolates in the cohort. Both ST8196 isolates were identified from the 2014 sample pool and displayed resistance to multiple antibiotics including carriage of $bla_{\text{CTX-M-15}}$ and chromosomal *parC* and *gyrB* mutations known to confer fluoroquinolone resistance (Table 1). In addition, the ST8196 isolates have acquired a plasmid-associated *qnrB4* gene. The isolation of ST8196 with plasmid-encoded *qnrB4* genes indicates that active surveillance for this ST is needed in Sydney and Australia more broadly.

In conclusion, MDR isolates in this genetically diverse BSI cohort exhibit limited variation in class 1 integron structure and in the cassette arrays they carry. IS*26* is playing a significant role in shaping the composition of genes in the CRL that include class 1 integrons in *E. coli* causing BSI. The role played by IS*26* in the evolution of CRL is significant because there are opportunities to acquire further antibiotic resistance gene cargo, including genes encoding resistance to clinically important antibiotics. We have identified *E. coli* from BSI that carry the plasmid-mediated fluoroquinolone resistance gene *qnrB4* and $bla_{\text{CTX-M-15}}$. These acquisition events may contribute to the rapid rise of co-resistance to these two clinically important antibiotics. Finally, IncF plasmids are important in the evolution of resistance and virulence

attributes in *E. coli* causing BSI in Australia. A detailed investigation of the plasmids that circulate in *E. coli* causing bacteraemia is warranted.

4.6 Acknowledgements and funding:

We wish to acknowledge Michael Liu for assistance with NextEra Sequencing Library preparation, and Cameron Reid for helping with bioinformatics workflow used in this manuscript.

This project was funded by the Australian Centre for Genomic Epidemiological Microbiology (AusGEM), a strategic partnership between the NSW-Department of Primary Industries and the University of Technology Sydney. PH is the recipient of the University of Technology Sydney PhD scholarship as a part of the AusGEM programme and the University of Technology International Research Scholarship (IRS). These results are helping to inform the OUTBREAK decision support system funded by the Medical Research Future Fund Frontier Health and Medical Research Program.

Transparency declarations

None to declare.

Published Manuscript:

https://doi.org/10.1099/mgen.0.000371

References:

1. Diekema DJ, Hsueh PR, Mendes RE, Pfaller MA, Rolston KV et al. The Microbiology of Bloodstream Infection: 20-Year Trends from the SENTRY Antimicrobial Surveillance Program. *Antimicrob Agents Chemother* 2019;63(7).
2. Institute TG. Stopping Sepsis A National Action Plan. Sydney, Australia: Australian Sepsis Network 2017.
3. Coombs G, Bell JM, Daley D, Collingnon P, Cooley L et al. Australian group on Antimicrobial Resistance Sepsis Outcomes Programs: 2017 Report. Report. Sydney2019. Report No.: Sydney: ACSQHC; 2019.
4. Turnidge J, Gottlieb T, Bell JM. Enterobacteriaceae Sepsis Outcome Programme (EnSOP) 2013 Antimicrobial Susceptibility Report2014.
5. Ananias M, Yano T. Serogroups and virulence genotypes of *Escherichia coli* isolated from patients with sepsis. *Braz J Med Biol Res* 2008;41(10):877-883.
6. Johnson JR. Virulence factors in *Escherichia coli* urinary tract infection. *Clin Microbiol Rev* 1991;4(1):80-128.
7. Coombs G, Bell JM, Daley D, Collingnon P, Cooley L et al. Australian Group on Antimicrobial Resistance. Sepsis Outcome Programs 2016 Report. . Sydney2018 Contract No.: Sydney: ACSQHC; 2018
8. Abernethy J, Guy R, Sheridan EA, Hopkins S, Kiernan M et al. Epidemiology of *Escherichia coli* bacteraemia in England: results of an enhanced sentinel surveillance programme. *J Hosp Infect* 2017;95(4):365-375.
9. Gillings M, Boucher Y, Labbate M, Holmes A, Krishnan S et al. The evolution of class 1 integrons and the rise of antibiotic resistance. *J Bacteriol* 2008;190(14):5095-5100.
10. Hall RM, Stokes HW. Integrons: novel DNA elements which capture genes by site-specific recombination. *Genetica* 1993;90(2-3):115-132.
11. Gundogdu A, Long YB, Vollmerhausen TL, Katouli M. Antimicrobial resistance and distribution of sul genes and integron-associated intI genes among uropathogenic *Escherichia coli* in Queensland, Australia. *J Med Microbiol* 2011;60(Pt 11):1633-1642.
12. Leverstein-van Hall MA, HE MB, AR TD, Paauw A, Fluit AC et al. Multidrug resistance among Enterobacteriaceae is strongly associated with the presence of integrons and is independent of species or isolate origin. *J Infect Dis* 2003;187(2):251-259.

13. Cummins ML, Reid CJ, Roy Chowdhury P, Bushell RN, Esbert N et al. Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene. *Microb Genom* 2019;5(2).

14. Djordjevic SP, Stokes HW, Roy Chowdhury P. Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Front Microbiol* 2013;4:86.

15. Reid CJ, Wyrsch ER, Roy Chowdhury P, Zingali T, Liu M et al. Porcine commensal *Escherichia coli*: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 2017;3(12).

16. Gillings MR. Integrons: past, present, and future. *Microbiol Mol Biol Rev* 2014;78(2):257-277.

17. Copur-Cicek A, Ozgumus OB, Saral A, Sandalli C. Antimicrobial resistance patterns and integron carriage of *Escherichia coli* isolates causing community-acquired infections in Turkey. *Ann Lab Med* 2014;34(2):139-144.

18. Horcajada JP, Soto S, Gajewski A, Smithson A, Jimenez de Anta MT et al. Quinolone-resistant uropathogenic *Escherichia coli* strains from phylogenetic group B2 have fewer virulence factors than their susceptible counterparts. *J Clin Microbiol* 2005;43(6):2962-2964.

19. Johnson JR, Kuskowski MA, O'Bryan TT, Maslow JN. Epidemiological correlates of virulence genotype and phylogenetic background among *Escherichia coli* blood isolates from adults with diverse-source bacteremia. *J Infect Dis* 2002;185(10):1439-1447.

20. Johnson JR, Kuskowski MA, Owens K, Gajewski A, Winokur PL. Phylogenetic origin and virulence genotype in relation to resistance to fluoroquinolones and/or extended-spectrum cephalosporins and cephamycins among *Escherichia coli* isolates from animals and humans. *J Infect Dis* 2003;188(5):759-768.

21. Moreno E, Prats G, Sabate M, Perez T, Johnson JR et al. Quinolone, fluoroquinolone and trimethoprim/sulfamethoxazole resistance in relation to virulence determinants and phylogenetic background among uropathogenic *Escherichia coli*. *J Antimicrob Chemother* 2006;57(2):204-211.

22. Carattoli A. Plasmids and the spread of resistance. *Int J Med Microbiol* 2013;303(6-7):298-304.

23. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* 2018;73(5):1121-1137.

24. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M et al. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J Bacteriol* 2010;192(21):5822-5831.

25. Gaastra W, Svennerholm AM. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol* 1996;4(11):444-452.

26. Pasqua M, Michelacci V, Di Martino ML, Tozzoli R, Grossi M et al. The Intriguing Evolutionary Journey of Enteroinvasive *E. coli* (EIEC) toward Pathogenicity. *Front Microbiol* 2017;8:2390.

27. Croxall G, Hale J, Weston V, Manning G, Cheetham P et al. Molecular epidemiology of extraintestinal pathogenic *Escherichia coli* isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. *J Antimicrob Chemother* 2011;66(11):2501-2508.

28. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin Microbiol Infect* 2014;20(5):380-390.

29. Andersen PS, Stegger M, Aziz M, Contente-Cuomo T, Gibbons HS et al. Complete Genome Sequence of the Epidemic and Highly Virulent CTX-M-15-Producing H30-Rx Subclone of *Escherichia coli* ST131. *Genome Announc* 2013;1(6).

30. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B et al. The epidemic of extended-spectrum-beta-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* 2013;4(6):e00377-00313.

31. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* 2014;111(15):5694-5699.

32. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK et al. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* 2011;6(10):e26578.

33. Poirel L, Gniadkowski M, Nordmann P. Biochemical analysis of the ceftazidime-hydrolysing extended-spectrum beta-lactamase CTX-M-15 and of its structurally related beta-lactamase CTX-M-3. *J Antimicrob Chemother* 2002;50(6):1031-1034.

34. Poirel L, Kampfer P, Nordmann P. Chromosome-encoded Ambler class A beta-lactamase of *Kluyvera georgiana*, a probable progenitor of a subgroup of CTX-M extended-spectrum beta-lactamases. *Antimicrob Agents Chemother* 2002;46(12):4038-4040.

35. Shaikh S, Fatima J, Shakil S, Rizvi SM, Kamal MA. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. *Saudi J Biol Sci* 2015;22(1):90-101.

36. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M beta-lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother* 2017;72(8):2145-2155.

37. Hwang JH, Shin GW, Hwang JH, Lee CS. Bloodstream Infection Due to CTX-M-15 and TEM-1 Extended-Spectrum beta-Lactamase-Producing *Salmonella enterica* serovar Virchow ST16. *Jpn J Infect Dis* 2017;70(3):308-310.

38. Paterson DL, Bonomo RA. Extended-spectrum beta-lactamases: a clinical update. *Clin Microbiol Rev* 2005;18(4):657-686.

39. Paterson DL, Hujer KM, Hujer AM, Yeiser B, Bonomo MD et al. Extended-spectrum beta-lactamases in *Klebsiella pneumoniae* bloodstream isolates from seven countries: dominance and widespread prevalence of SHV- and CTX-M-type beta-lactamases. *Antimicrob Agents Chemother* 2003;47(11):3554-3560.

40. Third Australian report on antimicrobial use and resistance in human health- AURA 2019-May 2019 [database on the Internet]2019. Available from: https://www.safetyandquality.gov.au/our-work/antimicrobial-resistance/antimicrobial-use-and-resistance-australia-surveillance-system-aura/aura-2019.

41. National Alert System for Critical Antimicrobial Resistances (CARAlert) [database on the Internet]2018. Available from: https://www.safetyandquality.gov.au/our-work/antimicrobial-resistance/antimicrobial-use-and-resistance-australia-surveillance-system-aura/national-alert-system-critical-antimicrobial-resistances-caralert.

42. Bell JM, Turnidge JD, Coombs GW, Daley DA, Gottlieb T et al. Australian Group on Antimicrobial Resistance Australian Enterobacteriaceae Sepsis Outcome Programme, Annual Report 2014. Australia: The Department of Health2016 Contract No.: 2.

43. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* 2012;18(3):268-281.

44. Darling AE, McKinnon J, Worden P, Santos J, Charles IG et al. A draft genome of Escherichia coli sequence type 127 strain 2009-46. *Gut Pathog* 2014;6:32.

45. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015;31(4):587-589.

46. Brettin et al. RastTK. *Scientific Reports* 2015. 8365.

47. Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3(10):e000131.

48. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.

49. Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014;2:e243.

50. Wickham H, Sievert C. *ggplot 2*: : *Elegant graphics for data analysis* 2016: New York Springer

51. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 2013;5(1):58-65.

52. Vangchhia B, Abraham S, Bell JM, Collignon P, Gibson JS et al. Phylogenetic diversity, antimicrobial susceptibility and virulence characteristics of phylogroup F *Escherichia coli* in Australia. *Microbiology* 2016;162(11):1904-1912.

53. Santos AC, Zidko AC, Pignatari AC, Silva RM. Assessing the diversity of the virulence potential of *Escherichia coli* isolated from bacteremia in Sao Paulo, Brazil. *Braz J Med Biol Res* 2013;46(11):968-973.

54. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis. *Int J Antimicrob Agents* 2018;52(3):430-435.

55. Moran RA, Hall RM. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. *Microb Drug Resist* 2018;24(4):411-421.

56. Daga AP, Koga VL, Soncini JGM, de Matos CM, Perugini MRE et al. *Escherichia coli* Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. *Front Cell Infect Microbiol* 2019;9:191.

57. Day MJ, Doumith M, Abernethy J, Hope R, Reynolds R et al. Population structure of *Escherichia coli* causing bacteraemia in the UK and Ireland between 2001 and 2010. *J Antimicrob Chemother* 2016;71(8):2139-2142.

58. Coque TM, Novais A, Carattoli A, Poirel L, Pitout J et al. Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerg Infect Dis* 2008;14(2):195-200.

59. Kallonen et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. 2017.

60. Bogema DR, McKinnon J, Liu M, Hitchick N, Miller N et al. Whole-genome analysis of extraintestinal *Escherichia coli* sequence type 73 from a single hospital over a 2 year period identified different circulating clonal groups. *Microb Genom* 2019.

61. Johnson TJ, Siek KE, Johnson SJ, Nolan LK. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J Bacteriol* 2006;188(2):745-758.

62. Johnson TJ, Wannemuehler Y, Johnson SJ, Stell AL, Doetkott C et al. Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. *Appl Environ Microbiol* 2008;74(22):7043-7050.

63. Peigne C, Bidet P, Mahjoub-Messai F, Plainvert C, Barbe V et al. The plasmid of *Escherichia coli* strain S88 (O45:K1:H7) that causes neonatal meningitis is closely related to avian pathogenic E. coli plasmids and is associated with high-level bacteremia in a neonatal rat meningitis model. *Infect Immun* 2009;77(6):2272-2284.

64. Ron EZ. Host specificity of septicemic *Escherichia coli*: human and avian pathogens. *Curr Opin Microbiol* 2006;9(1):28-32.

65. Reid CJ, McKinnon J, Djordjevic SP. Clonal ST131-H22 *Escherichia coli* strains from a healthy pig and a human urinary tract infection carry highly similar resistance and virulence plasmids. *Microb Genom* 2019;5(9).

66. Skyberg JA, Johnson TJ, Johnson JR, Clabots C, Logue CM et al. Acquisition of avian pathogenic *Escherichia coli* plasmids by a commensal *E. coli* isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. *Infect Immun* 2006;74(11):6287-6292.

67. Zhu Ge X, Jiang J, Pan Z, Hu L, Wang S et al. Comparative genomic analysis shows that avian pathogenic *Escherichia coli* isolate IMT5155 (O2:K1:H5; ST complex 95, ST140) shares close relationship with ST95 APEC O1:K1 and human ExPEC O18:K1 strains. *PLoS One* 2014;9(11):e112048.

68. Partridge SR, Tsafnat G, Coiera E, Iredell JR. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol Rev* 2009;33(4):757-784.

69. An XL, Chen QL, Zhu D, Zhu YG, Gillings MR et al. Impact of Wastewater Treatment on the Prevalence of Integrons and the Genetic Diversity of Integron Gene Cassettes. *Appl Environ Microbiol* 2018;84(9).

70. Reid CJ, Roy Chowdhury P, Djordjevic SP. Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. *Plasmid* 2015;80:127-137.

71. Roy Chowdhury P, Charles IG, Djordjevic SP. A role for Tn6029 in the evolution of the complex antibiotic resistance gene loci in genomic island 3 in enteroaggregative hemorrhagic *Escherichia coli* O104:H4. *PLoS One* 2015;10(2):e0115781.

72. Roy Chowdhury P, Ingold A, Vanegas N, Martinez E, Merlino J et al. Dissemination of multiple drug resistance genes by class 1 integrons in *Klebsiella pneumoniae* isolates from four countries: a comparative study. *Antimicrob Agents Chemother* 2011;55(7):3140-3149.

73. Roy Chowdhury P, McKinnon J, Liu M, Djordjevic SP. Multidrug Resistant Uropathogenic *Escherichia coli* ST405 With a Novel, Composite IS26 Transposon in a Unique Chromosomal Location. *Front Microbiol* 2018;9:3212.

74. Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP et al. Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in *Escherichia coli* of human and animal origin. *PLoS One* 2010;5(9):e12754.

75. Gillings MR. Class 1 integrons as invasive species. *Curr Opin Microbiol* 2017;38:10-15.

76. Harmer CJ, Hall RM. IS*26*-Mediated Precise Excision of the IS26-aphA1a Translocatable Unit. *MBio* 2015;6(6):e01866-01815.

4.8 Figures, tables and supplementary data:

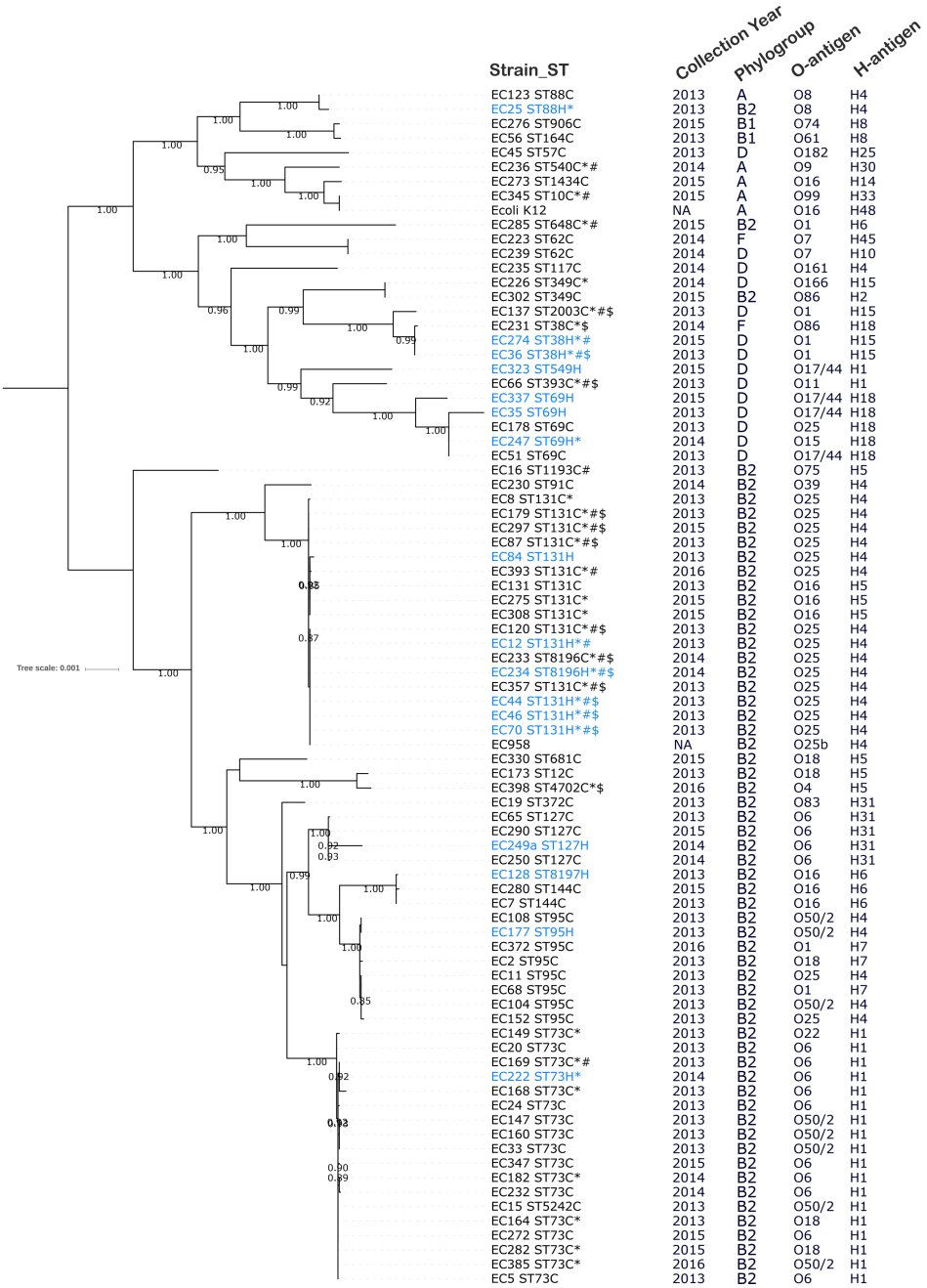


Figure 1:A mid-point rooted, maximum likelihood phylogenetic tree of isolates included in the cohort. The panels on the right-hand side indicate their year of isolation, *in-silico* phylo-grouping, o-antigen and H-antigen profiles. The tree includes two reference genomes, *E. coli* K12 and EC958. The strain labels coloured in blue indicate isolates from hospital acquired infections. MDR strains are marked with a *, ESBL producers with $ and fluoroquinolone resistant isolates with #. Year of isolation for the two

reference genomes were not available and is hence labelled as NA in the metadata column.

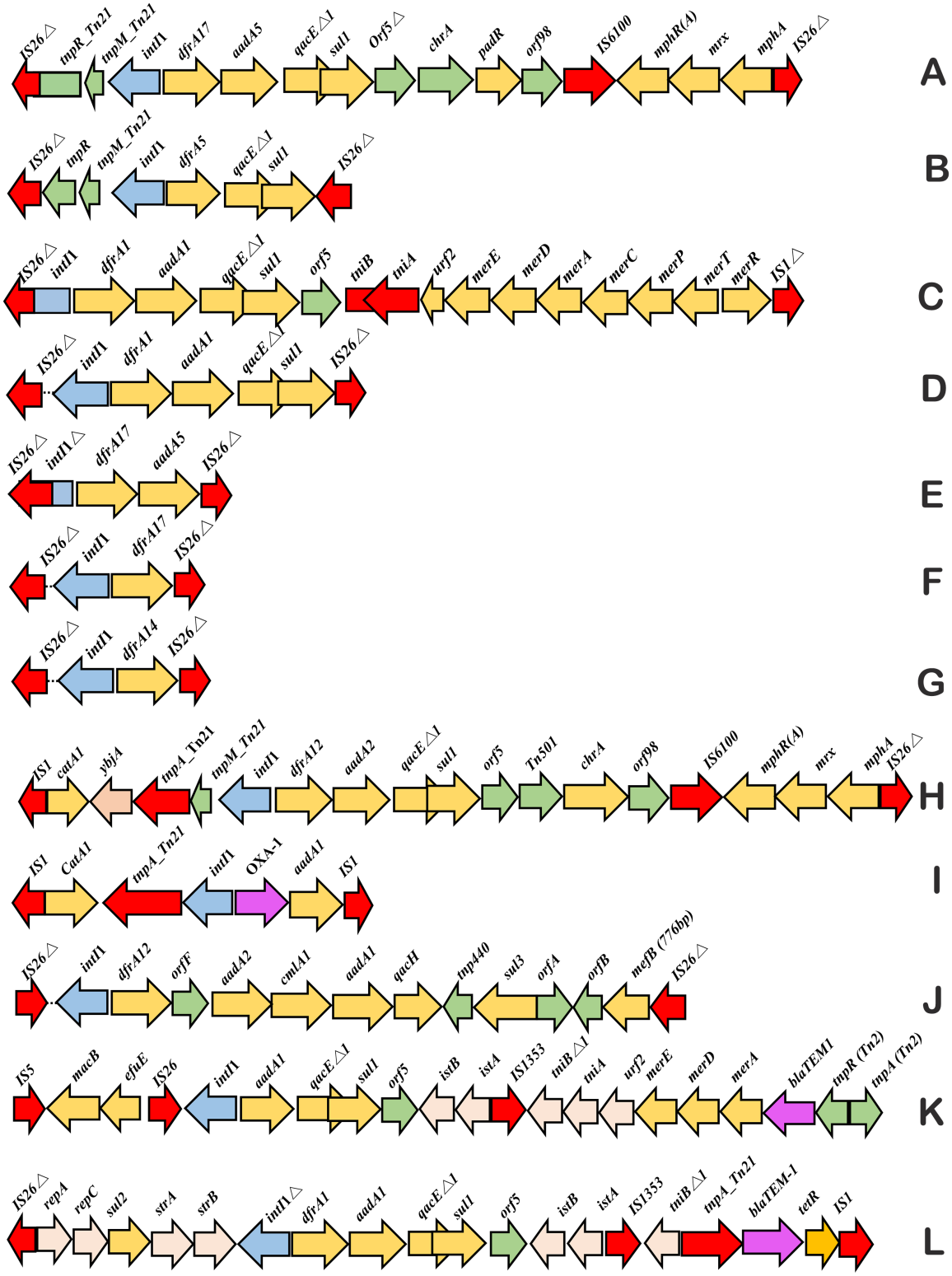


Figure 2: Diagrammatic representation of class 1 integron associated complex resistance loci identified in this study. The terminal IS*26*s indicated as red arrows are mostly fragmented, and their relative orientation indicated in the diagram is based on

analysis of the partial IS*26* sequence. Other IS elements also indicated as red arrows. The antimicrobial resistance genes are indicated as yellow arrows and overlap when fragmented. Class 1 integron (full or fragmented) indicated as blue color arrows and ESBL encoding genes depicted as pink arrows. All orfs and *tnp* genes colored as green arrows. The *rep* genes and other genes indicated as light pink arrows. Figure not to scale.

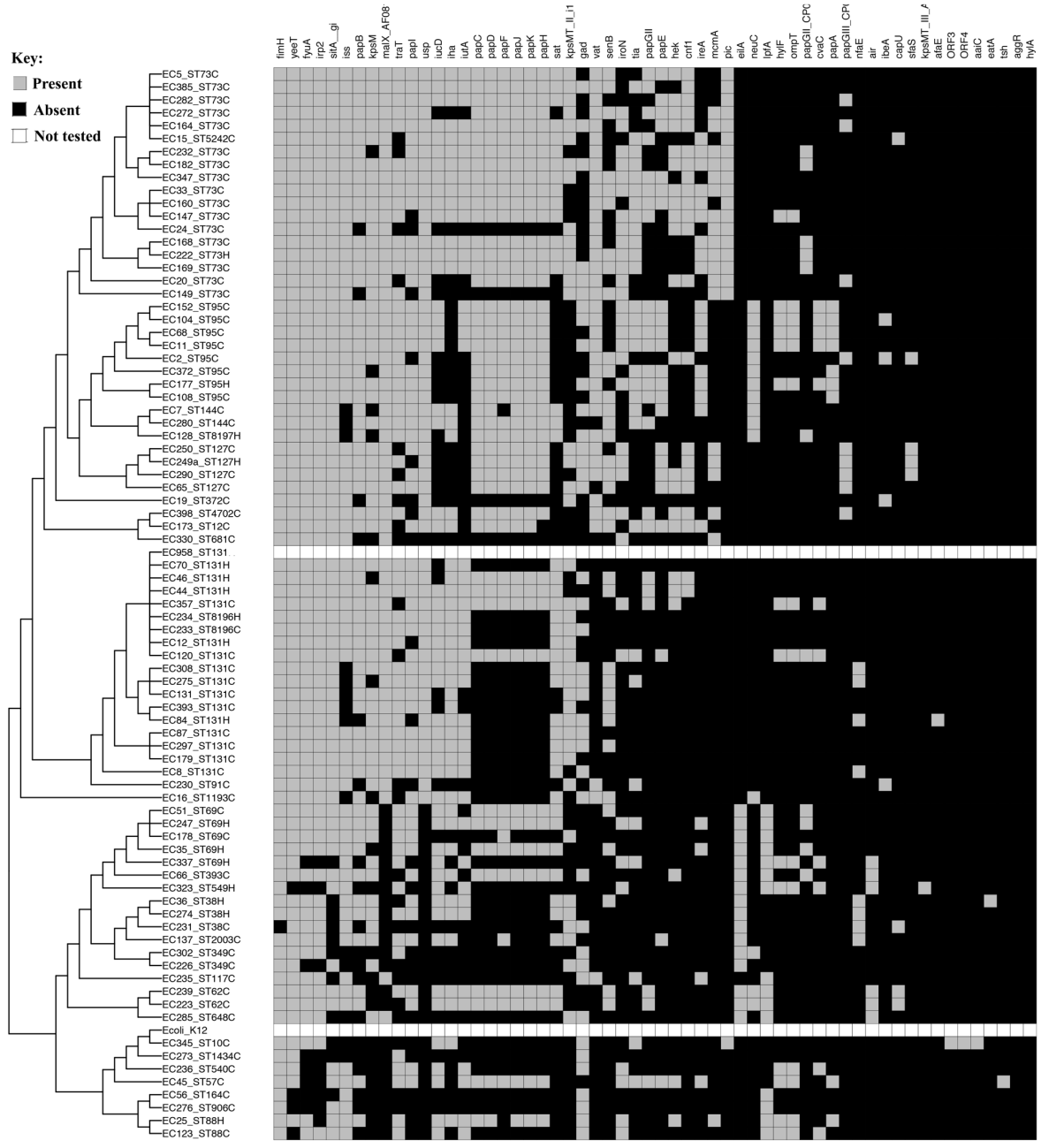


Figure 3: A presence/absence heat-map of *E. coli* virulence genes. The phylogenetic relationship of the isolates is presented as a cladogram along the Y axis and the virulence genes are listed in the top panel along the X-axis. Grey box indicates presence and black box indicates absence of a gene. Virulence gene profiling was not done for the

two reference genomes (EC958 and *E. coli* K12), hence the corresponding rows are left blank.

Table1_Chapter4: Resistance genes and plasmid profile of isolates sequenced and analysed in this study. The isolate descriptor column consists of isolate number, the sequence type (ST) and Hospital Onset (H) or community onset (C) included as suffix, followed by symbols to include phenotypic profile, where MDR is indicated as *, ESBL producers indicated as $ and fluoroquinolone resistant isolates indicated as # (find attached separate word file named as Table1_Chapter4).

S1_Chapter4: Clinical metadata sheet for isolates included in this project. Abbreviations listed in the primary source of infections column stand for: UTI= urinary tract infection including pyelonephritis, BTI= biliary tract infection including cholangitis, IAI= intra-abdominal infection other than biliary tract, NF= no focus or febrile neutropenia and infections, SS= Skin and skin structure and N/A = data not available (find attached separate excel sheet file named as S1_Chapter4).

S2_Chapter4: Assembly statistics of the genomes sequenced in this study and their Genbank accession numbers (find attached separate excel sheet file named as S2_Chapter4).

Table S3_Chapter4: List of genes used for virulence profiling of isolates, function of proteins encoded by the genes and number of isolates that harbour the virulence gene (find attached separate word file named as TableS3_Chapter4).

Table S4_Chapter4: Virulence intensity of isolate genomes assessed by the total number of genes of the 58 genes included in the ARIBA database, their phylogroup and predicted disease onset location (find attached separate word file named as TableS4_Chapter4).

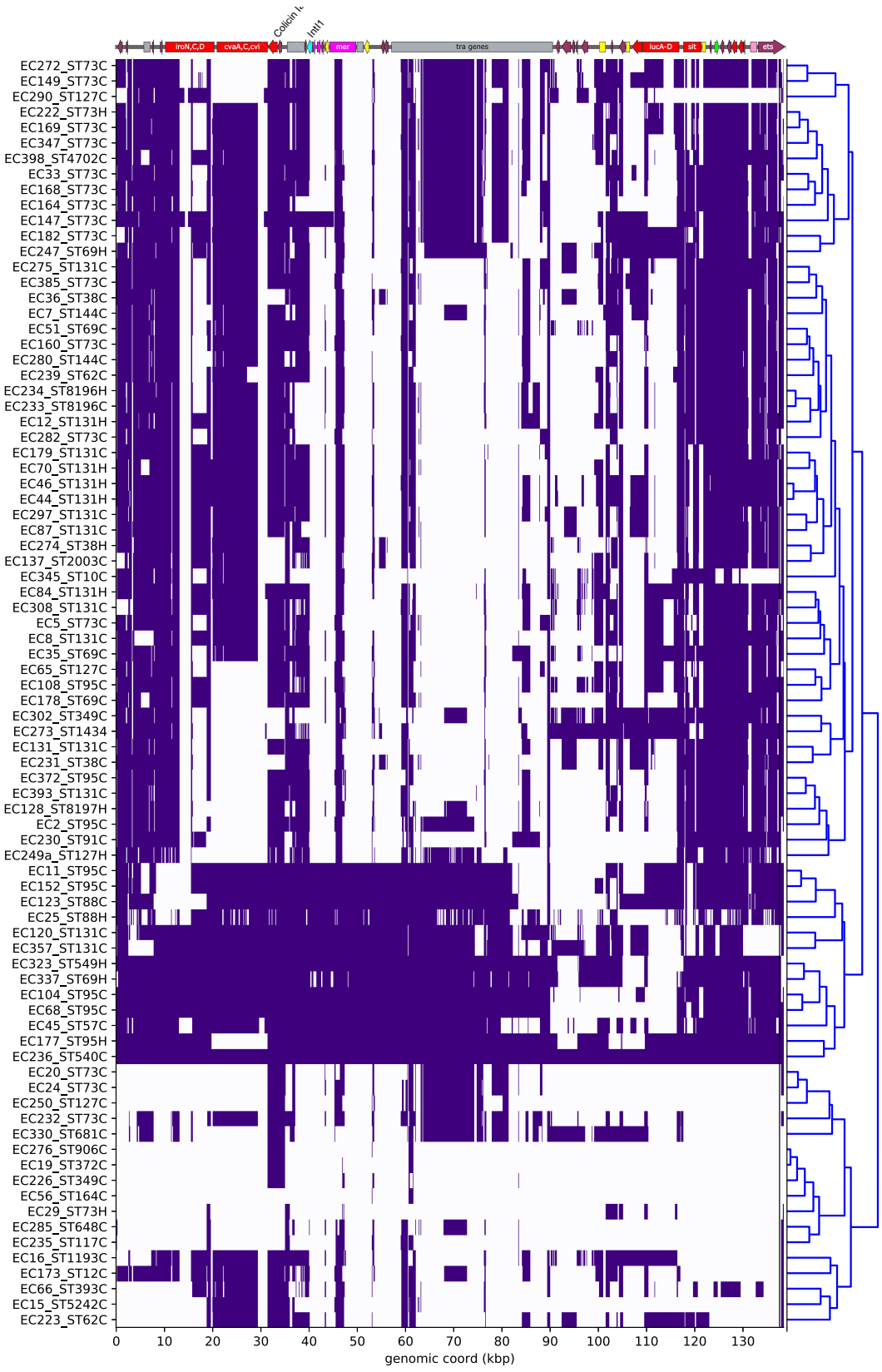


Figure S5: Illumina raw read mapping of genomes with plasmid pCERC4 (accession no. KU578032). The purple colour indicates the >95% identity of that region to the mapped plasmid and white indicates no match according to BLASTn alignment.

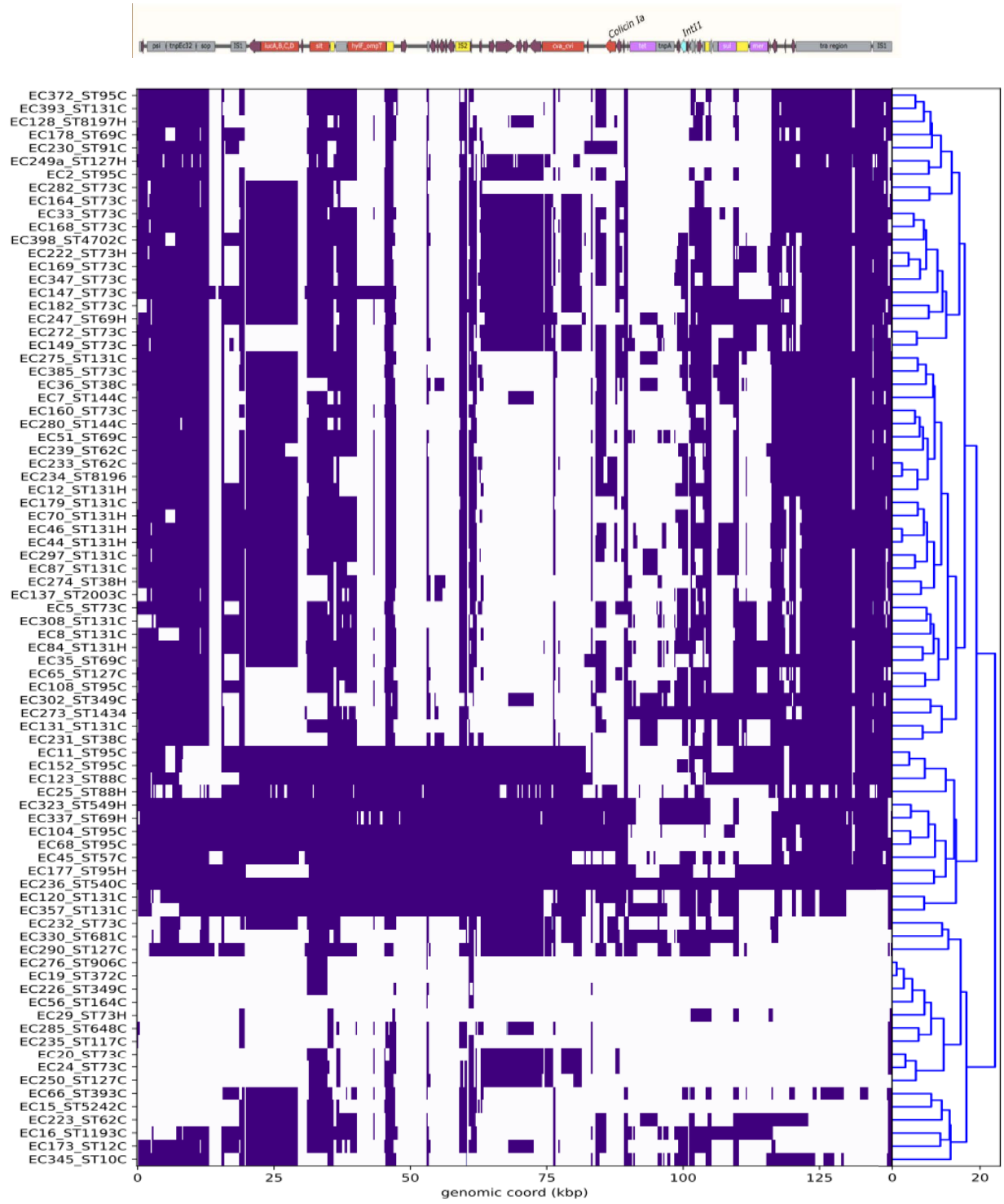


Figure S5: Cladogram of Illumina raw read mapping of genomes with plasmid pSDJ2009-52F (accession no. MH195200.1). The purple colour indicates the >95% identity of that region to the mapped plasmid and white indicates no match according to BLASTn alignment.

Chapter 5: *Escherichia coli* ST8196 is a novel, locally evolved member of the ST131 clonal complex

Priyanka Hastak[1], Mathieu Fourment[1], Aaron E. Darling[1], Thomas Gottlieb[3,4], Elaine Cheong[3], John Merlino[3,4], Garry S. A. Myers[1], Steven P. Djordjevic[1,2] and Piklu Roy Chowdhury[1,2*]

[1]The ithree institute, University of Technology Sydney, City Campus, Ultimo, NSW 2007, Australia.

[2]NSW Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Sydney, Australia.

[3]Department of Microbiology and Infectious Diseases, Concord Hospital, NSW Health Pathology, Hospital Road, Concord 2139, NSW, Australia.

[4] Faculty of Medicine, University of Sydney, NSW Australia.

Abstract:

The H30Rx subclade of Escherichia coli ST131 is a clinically important, globally dispersed pathogenic lineage that typically displays resistance to fluoroquinolones and extended spectrum β-lactams. Isolates EC233 and EC234, variants of ST131-H30Rx with a novel sequence type (ST) 8196, isolated from unrelated patients presenting with bacteraemia at a Sydney Hospital in 2014 are characterised here. EC233 and EC234 are phylogroup B2, serotype O25:H4A, and resistant to ampicillin, amoxicillin, cefoxitin, ceftazidime, ceftriaxone, ciprofloxacin, norfloxacin and gentamicin and are likely clonal. Both harbour an IncFII_2 plasmid (pSPRC_Ec234-FII) that carries most of the resistance genes on an IS26 associated translocatable unit, two small plasmids and a novel IncI1 plasmid (pSPRC_Ec234-I). SNP-based phylogenetic analysis of the core genome of representatives within the ST131 clonal complex places both isolates in a subclade with three clinical Australian ST131-H30Rx clade-C isolates. A MrBayes phylogeny analysis of EC233 and EC234 indicates that these strains belonging to ST8196 share a most recent common ancestor with ST131-H30Rx strain EC70 isolated from the same hospital in 2013. Our study identified genomic hallmarks that define the ST131-H30Rx subclade in the ST8196 isolates and highlights a need for

unbiased genomic surveillance approaches to identify novel high-risk MDR E. coli pathogens that impact healthcare facilities.

5.1 Introduction:

Escherichia coli sequence type (ST) 131 is a globally dominant pandemic multi-drug resistant (MDR) lineage that causes community-acquired (CA) and hospital-acquired (HA) blood stream infections (BSIs). It is most frequently associated with urinary tract infections (UTIs) and BSIs in developed and developing countries. Its association with the dissemination of fluroquinolone and cephalosporin resistance genes, causing major multi-drug resistant pandemics infections world-wide [1], is well documented. The MDR lineage is not limited to clinical infections and has been isolated from a variety of sources such as food producing animals, mammals, birds, cats, dogs, sheep, agriculture/horticulture, water and soil [2] [3].

Eighty-five sequence types form the ST131 clonal cluster. The majority are isolates collected from MDR human infections (https://enterobase.warwick.ac.uk/). In this study, we characterise a novel sequence type, ST8196, that was designated as a member of the ST131 clonal cluster. Two ST8196 isolates, denoted EC233 and EC234, were collected one day apart from two different patients presenting bacteraemia at Concord Repatriation Hospital in Sydney in November 2014. Both isolates were resistant to third generation amino-penicillins (ampicillin and amoxicillin), second and third generation cephalosporins (cefoxitin, ceftazidime and ceftriaxone), quinolones (ciprofloxacin and norfloxacin), and the aminoglycoside antibiotic gentamycin. They were classified as multi-drug resistant (MDR) [4]. Based on a 48-hour post-admission cut-off criterion for sample collection, EC233 was from a CA infection, while EC234 was from a HA infection.

The primary objective of this study was to conduct molecular characterisation two ST8196 isolates, EC233 and EC234 that belong to ST131CC. We present a comprehensive characterisation of the ST8196 genomes and, using phylogenetic analysis, we propose that ST8196 has evolved from MDR ST131 clones circulating in Australia. This study also highlights the ongoing evolution, persistence and clustering of MDR within novel and established *E. coli* plasmid lineages circulating within ST131 isolates that are causing BSIs in Australia.

5.2 Methodology:

5.2.1 Bacterial strains and growth conditions:

E. coli EC233 and EC234 were sub-cultured in the lab on Lysogeny Broth (LB) agar plates or broth at 37°C for 18 hours. Genomic DNA was extracted from overnight broth cultures grown with aeration on a rotary shaker set at 2,500 rpm. Antibiotic susceptibility profiles were generated using the VITEK2 (BioMerieux) system and interpreted using the European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoints [5]. Synergy tests were used to phenotypically confirm production of extended spectrum beta lactamase (ESBL) enzymes.

5.2.2 Genomic DNA extraction, library preparation and whole genome sequencing:

For Illumina HiSeq sequencing, genomic DNA was extracted using 2ml of overnight culture with the ISOLATE II Genomic DNA kit (Bioline, Australia), following the manufacturers protocol, and quantified using the Qubit fluorimeter and dsDNA HS Assay kit (Thermo Fisher Scientific, Australia). Whole genome sequencing libraries were prepared from 2ng of gDNA using the Illumina Nextera kit [20]. The Illumina HiSeq 2500 v4 sequencer (Illumina, San Diego, CA, USA) in rapid PE150 mode was used to sequence these libraries. The assembled genomes were submitted in GenBank under the accession number RDUX00000000 and RDUW00000000.

Isolate EC234 was also sequenced using the PacBio Single Molecule Real Time (SMRT) sequencing platform. Genomic DNA for SMRT sequencing was isolated from 1.8ml of overnight cultures using a moBio genomic DNA kit (Qiagen, Germany), following the manufacturers protocol. Sequencing was performed at the Ramaciotti Center for Genomics, University of New South Wales (Sydney, Australia). Demultiplexed PacBio long reads were assembled with the Illumina short reads, using the Unicycler [6] hybrid genome assembly protocol. The assembled genome was submitted in GenBank under accession number WMLE00000000.

5.2.3 Genotypic profiling and genome-wide comparison of isolates:

The eMLST, Resfinder, virulence finder database (VFDB), PlasmidFinder, serotype finder and CHtyping databases were used to identify resistance and virulence genes, the plasmid types

and serotype analysis (https://cge.cbs.dtu.dk/services/) (Figure 2B). New sequence type designations were acquired from the Enterobase database (http://enterobase.warwick.ac.uk/species/index/ecoli). Average nucleotide identity was calculated using the online ANI calculator (http://enve-omics.ce.gatech.edu/ani/newjob). Preliminary genome annotations were generated using an online version of RASTtk [7]. Putative antimicrobial resistance genes of interest were confirmed using stand-alone BLASTn analyses and an in-house database at UTS, prior to manual verification using NCBI-ORF finder. Iterative BLASTn and BLASTp searches [8] were used for confirmation of annotated plasmids. Only legitimate ORFs showing >95% sequence similarity (E-value 0.001) across 100% of the query sequence were considered for further analysis. Figures comparing BLASTn alignment of genome and plasmid sequences were generated using BRIG v0.95 [9] set at 90% lower and 100% upper identity threshold, SnapGene v3.3.4 (GSL Biotech) and EasyFig version 2.2.2 [10]. Roary [11] was used for Pangenome analysis.

5.2.4 Phylogeny analysis of ST8196 isolates:

The SNP phylogenies of ST8196 genomes and plasmids were inferred with parSNP, using the -c and -x flags to evoke forced alignment across large co-linear blocks and recombination filtering to improve clade confidence scores [12]. *E. coli* EC958 ST131 (NZ_HG941718.1) was used as a reference genome for all ST131 phylogenies. The Bayesian Markov chain Monte Carlo method as implemented in the MrBayes package (version 3.2.7a) was used on 148 genomes aligned using parSNP.

We utilized the GTR substitution model with the gamma model of heterogeneity (4 categories). We placed an unconstrained compound gamma-Dirichlet prior on the branch lengths for the tree and ran the chain for 100 million iterations. Phylogenetic tree figures were generated using FigTree2 (http://tree.bio.ed.ac.uk/software/figtree/) and compiled with metadata online version of iToL (https://itol.embl.de/) [13].

5.3 Results and Discussions:

5.3.1 EC233 and EC234 represented ST8196, a novel sequence type within ST131 clonal complex:

In silico multi-locus sequence typing (MLST) of the genomes of two isolates (EC233 and EC234) assembled from Illumina short read sequences led to the establishment of a novel

sequence type, ST8196 (phylogroup B2, serotype H4:O25), within the ST131 clonal cluster in Enterobase repository. Of the 7 genes used for Achtman *E. coli* MLST profiling, ST131 is characterised by allelic variant 13 of the *icd* gene encoding iso-citrate dehydrogenase, while ST8196 had allelic variant 912. A whole-genome SNP based phylogeny analysis (**Figure 1 from Chapter 4**) using EC958 as the ST131 reference genome, investigated the relatedness of ST8196 with other genomes representing 85 different STs within the ST131 clonal cluster. We also included all ST131 isolates in our collection and, based on alignment of 73% of the core genome, the nearest common ancestor to ST8196 is EC70, an ST131 isolate in our collection.

The two ST8196 isolates, EC233 and EC234, and ST131 isolate, EC70 clustered into a subclade with ESC_RA8813AA, a ST8189 isolate. Based on the MLST profiles, ST8189 differed from ST131 and ST8196 in the *mdh* gene (alleleic variant number 697), encoding malate dehydrogenase. The ST8189 isolate included in the SNP tree was collected in 2017, however there are three other representatives of ST8189 in the Enterobase database, including an isolate collected in 2017 from Australia. The assembled genome of the other three representatives was not available at the time of analysis and was not included in our analyses. Of the core genome considered in the SNP phylogeny analysis, there were only 2 SNP differences between the two ST8196 isolates, EC233 and EC234. The mean average nucleotide identity between the two genomes was 99.99%, indicative of clonality. There were 17 SNP differences between the ESC_RA8813AA_ST8189 and the EC234_ ST8196, and 12 SNP differences between EC70_ST131 and EC234_ST8196.

As the most recent common ancestor to EC233 and EC234 was EC70, an ST131 isolate in our collection from sampling year 2013, we embarked upon a time-resolved genealogy analysis of the ST8196 isolates using a global collection of ST131s. Our objective was to assess whether ST8196 evolved locally in Australia or, if there were other ST131 genomes closely related to EC70, suggesting that ST8196 could share a common ancestor with ST131. Tempest analysis (http://tree.bio.ed.ac.uk/software/tempest/) of aligned regions from 653 ST131 genomes generated in the course of a parSNP phylogeny analysis (date and country of isolation retrieved from the Enterobase database, 30th May 2019,) indicated the correlation between time and substitutions is 0.12. MrBayes analysis on a subset of 158 genomes (**File S1_Chapter5, Figure S2**), including the widest geographical distribution of ST131 genomes in Enterobase, together with all isolates in our collection and the ST8196 genomes, identified a subclade of 8 ST131 genomes (**Figure S2 and 2A**). The subclade comprised of isolates from Australia and United

States collected between 2011 and 2015 from human blood or urine specimens. However, this subclade had internal branching, and our ST8196 isolates EC233 and EC234 clustered together into a smaller group with isolates from Australia, including EC70 (2013 isolate) and ESC_CA5436 (2011 isolate) that diverged from a common ancestor shared with an isolate from the USA, FA7284AA, collected in 2015 from human urine.

5.3.2 Genotypic profiling of ST131 and ST8196 isolates that cluster together suggest major differences are within accessory regions of the genomes:

All 8 isolates in the subclade (**Figure 2A**) had identical serotypes (H4:O25), fimH types (H30), CH types (40-30) and a common pool of virulence associated genes, including iha (encoding adherence protein), iss (increased serum survival gene, supporting immune evasion) and sat (secreted autotransporter toxin) (**Figure 2B**). Three isolates (EC234_ST8196, ESC_PA1114AA_AS, ESC_FA7278AA_AS) additionally had one or two copies of the immune suppression and survival gene, gad, encoding glutamate decarboxylase. Isolate ESC_FA7278AA_AS from the United States shared a closest common ancestor with the Australian ST131 and ST8196 isolates and possesses cnf1 (cytotoxic necrotising factor) and the senB (plasmid encoded enterotoxin) genes. The major differences between the clustered ST8196 and ST131 genomes were found in resistance determinants and plasmid content (**Figure 2B**), indicating the likely contribution of the accessory gene pool in the emergence of ST8196 genomes.

A pairwise BLASTn alignment (**Figure 2C**) of the closely related genomes Illumina assembled sequences present in the EC234_ST8196 clade, indicates >98% sequence identity across most of the chromosome (**Figure 2C**). A few regions have identity dropping below 90% in some genomes, indicated in purple (FA6284AA), green (GA1863AA) and blue (EC70_ST131) rings. Major differences were localised on unitigs beyond the 5MB region, which indicate alignments with the four closed plasmids, and 4 additional unitigs representing the EC234 genome. The innermost circle (grey) in **Figure 2C** represents alignment with EC233_ST8196, indicating nearly 100% identity over the entire genome including accessory regions. Yellow radiating bars along in the middle of the figure (**Figure 2C)** indicate positions of genomic

islands in EC234 (**S3_Chapter5**), identified by the island-viewer software (https://www.pathogenomics.sfu.ca/islandviewer/). All ST131 isolates included in the pairwise BLASTn analysis, irrespective of the country of isolation, revealed near identity across the genomic islands in EC234 indicating sequence conservation within the chromosomal regions of the isolates.

5.3.3 Pangenome analysis ST131 subclade with ST8196 isolates in Mr Bayes phylogeny analysis:

Roary [11] was used to identify and distinguish between the pangenome and accessory genome content of the isolates. Based on ≥95% amino acid sequence identity and MCL clustering of peptides encoded by open reading frames, 5853 clusters were identified in the pangenome and included in the analysis. The core genome of the isolates comprised of 4527 clusters (**Figure 2D**) and was present at least once in 7 of the 8 genomes. Gene clusters forming soft-core (i.e., present in 95% of the 99% of genomes included in the analysis) [14] were not detected, indicating strict conservation of genes within the core genome of ST131 and ST8196 isolates included in this analysis. The accessory genome could be divided into two subsets, gene clusters forming the "shell" (815) and the "cloud" (611). The "shell" represented the number of gene clusters present in at least 7 of the 8 the genomes, while the "cloud" comprised of unique clusters present in at least 1 of the genomes in the isolate pool. Presence and absence gene sets (File S4) from Roary analysis reinforced the prediction that the major differences in these genomes comprised of gene clusters that are typically influenced by lateral gene transfer events [15] [11].

A binary tree created from the presence and absence of gene clusters in the accessory gene pool demonstrated that the Australian ST131 genomes grouped with ST8196 genomes, in topological congruence with the SNP based phylogeny (**Figure 2A, Figure 1 and S2**). By correlating the pangenome data with genotypic profiling shows that the ST8196 isolates had three typeable plasmids: an IncFII plasmid, an IncI plasmid and a ColBS512-like plasmid. While the IncI plasmid was unique to the ST8196 isolates, the ColBS512-like plasmid was a distinguishing feature of the FA6284AA_AS_ST131 genome collected in 2015 from the United States. All other ST131 genomes were characterised by the presence of IncFIA, IncFII and IncFIB replicons, including EC70_ST131 from our collection. Additionally, isolate PA1114AA_AS, a ST131 isolate collected from Australia in 2014 had an IncX1 plasmid.

5.3.4 Plasmids in EC234_ST8196:

Unicycler hybrid assembly of long and short read sequences of EC234 resolved the genome into 9 unitigs, including a 5.09Mb unitig representing the major portion of the chromosome, 4 closed plasmids, and 4 additional unitigs which likely forms the remaining portions of the chromosome. Unitig 2 typed as a 92,955nt IncI plasmid, while unitig 3 was an 85,199nt long IncFII plasmid. The isolate also had two small plasmids, one represented by unitig 5, 5164nt long (46x coverage) consisting of a *repA* and *mobA* gene in addition to 3 hypothetical genes and had 99.7% identity over 100% query length with plasmid *E. coli* pEc631_5 (CP040268.1), collected from bivalve mollusc from marine specimen in Norway.

There are also several other *E. coli* plasmids with >99% identity over the query length in GenBank collected from other geographical regions. Unitig 6 was a 2,101nucleotide (112.8x coverage) plasmid, identical to a plasmid from *Shigella sonnei* Ss046 (pSS046_spC, CP000643.1), collected from a patient with bacillary dysentery in China. This plasmid types as a ColBS512 like plasmid, as the *rep* gene was identical to *S. boydii* CDC 3083-94 plasmid pBS512_2 (CP001058.1).

The IncFII plasmid encoded most genes accounting for the resistance phenotype **(Figure 3A)** of EC234 except the ESBL phenotype, which was most likely an effect of the $bla_{\text{CTX-M-15}}$ gene located on unitig 1, together with *gyrB* and *parC* mutations contributing to fluoroquinolone resistance. The plasmid however had a *qnrB4* gene capable of imparting species-specific levels of fluroquinolone resistance to different members of Enterobacteriales.

The resistance loci consisted of a fragmented class 1 integron with a *dfrA17-aadA5* gene cassette-array, although both the integrase gene (*intI1*) and the aminoglycoside gene (*aadA5*) are interrupted by IS*26.* There are two additional copies of IS*26* in the resistance loci flanking the aminoglycoside resistance gene *aac(3)-IIa.* Other features of the resistance loci included a remnant of the 3'-*CS* of a class I integron with an IS*6100* associated macrolide resistance module and the $bla_{\text{DHA-1}}$ beta-lactamase gene. As IncF replicons are abundant in ST131 genomes, we attempted to test for the presence of similar plasmids in ST131 isolates in our collection, and within closely related ST131 genomes identified in our phylogeny analysis **(Figure 3B)**. Only EC233 had an identical plasmid. The plasmid replication and stability genes present (represented with light green arrows in the plasmid map) in the ST8196 IncFII plasmid are missing in the other genomes. As such ST8196 isolates carry an IncFII_2 plasmid, variants

of which likely distinguish them from the population of IncFII plasmids circulating within E. coli ST131 in Australia

The incI plasmid in EC234 did not possess resistance genes, however it does have the full complement of conjugative (type IV pili) and transfer genes, indicating an ability to mobilise independently. We downloaded 500 completely closed IncI plasmids from the plasmid database [19] on 23rd September 2019, and selected a subset of 38 IncI plasmids from *E. coli, Shigella* and *Salmonella* spp. with >99% identity over >60% of the input query sequence (Unitig 2, EC234_ST8196) for a SNP phylogeny analysis **(Figure 4A)**. Based on the alignment of 59% of the plasmid backbone in all 38 plasmids, the IncI plasmid from EC234 was most closely related to two other IncI plasmids from *E. coli* ST131 isolates collected from urine samples in Australia (NZ_CP035469.1 and NZ_CP035721.1), and an *E. coli* plasmid (NZ_LS992187.1) reported from Germany.

Alignment of the plasmids showed that differences clustered around hypothetical genes on the plasmid backbone, further supporting that similar IncI plasmids are circulating in Australia. The two IncI plasmids from Australia (NZ_CP035469.1 and NZ_CP035721.1) may have evolved from a common ancestral ST131, raising the possibility of local lateral acquisition of the plasmid by the ST8196 clones EC233 and EC234.

5.4 Conclusions:

We present a detailed genomic characterisation of EC233 and EC234, representing the first description of ST1896 within the globally pandemic ST131 clonal cluster. Phylogenetic analyses suggest that ST8196 has most likely evolved independently from EC70_ST131 and is related to ST8189 within the ST131 clonal cluster. Isolates representing ST8189 have also been recently reported from Australia, however our isolates were more closely related to EC70_ST131 in our sample pool than ST8189 isolate. Genes conferring resistance to the extended spectrum beta-lactam and fluoroquinolone groups of antibiotics were chromosomally located in ST8196, with the ESBL gene, *bla*$_{\text{CTX-M-15}}$, encased in a laterally mobile module associated with IS*Ecp1*.The IncFII plasmid carried the remaining resistance cargo and was distinct from the IncF plasmids circulating in ST131 in our collection. The incI plasmid in ST8196 isolates is most likely a progenitor of other IncI plasmids circulating within ST131 clones in Australia.

The dissemination of MDR plasmids within clinical settings complicates treatment options. IncF plasmids are common drivers of MDR phenotypes within ST131 *E. coli*, almost exclusively carrying clinically important antimicrobial resistance genes and ESBLs. They are frequently reported to carry the CTX-M type of ESBLs that co-transfers plasmid mediated quinolone resistance and aminoglycoside resistance [16]. The CTX-M cluster 1 genes, including CTX-M-15 [17], is primarily disseminated via IncFII:IncFIA plasmids and sometimes on IncI plasmids, while the CTX-M cluster, including CTX-M-14 and CTX-M-27 are usually associated with IncHI2 plasmids and sometimes with IncFII [18]. Although the MDR gene cluster and the ESBL and fluoroquinolone resistance genes are not linked on IncFII or the IncI plasmids in ST8196, association of ISEcpI with the chromosomal CTX-M15 gene raises the possibility of mobilisation any point in time.

The proposed local evolution of this novel sequence type is based on an extensive SNP-based phylogenetic analysis of the core genome. This contention is supported by phylogenetic clustering of the accessory genome. The MDR ST8196 isolates characterised in this study are armed with virulence and antimicrobial resistant features that have contributed to the local establishment of these clones. With the potential for further lateral acquisitions, these clones have the potential to trigger disease with significantly curtailed treatment options.

Published Manuscript:

doi: 10.1080/22221751.2020.1797541

5.5 References:

1. Pitout JD. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. Front Microbiol. 2012 Jan 3(9):1–7.
2. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD, Gomes Moriel D, Peters KM, Davies M, Rogers BA, Dougan G, Rodriguez-Baño J, Pascual A, Pitout JD, Upton M, Paterson DL, Walsh TR, Schembri MA, Beatson SA. Global dissemination of a multidrug resistant *Escherichia coli* clone. Proc Natl Acad Sci U S A. 2014 Apr 15;111(15):5694-9. doi: 10.1073/pnas.1322678111. Epub 2014 Mar 31. PubMed PMID: 24706808; PubMed Central PMCID: PMC3992628.
3. Reid, C. J., Wyrsch, E. R., Roy Chowdhury, P., Zingali, T., Liu, M., Darling, A. E., … Djordjevic, S. P. (2017). Porcine commensal *Escherichia coli*: a reservoir for class 1 integrons associated with IS*26*. Microbial genomics, 3(12), e000143. doi:10.1099/mgen.0.000143

4. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, Harbarth S, Hindler JF, Kahlmeter G, Olsson-Liljequist B, Paterson DL, Rice LB, Stelling J, Struelens MJ, Vatopoulos A, Weber JT, Monnet DL. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect.* 2012 Mar;18(3):268-81. doi: 10.1111/j.1469-0691.2011.03570.x. Epub 2011 Jul 27. PubMed PMID: 21793988.
5. Bell, J.M., Turnidge, J.D., Coombs, G.W., Daley, D.A., Gottlieb, T., Robson, J. and George, N. (2014) Australian Group on Antimicrobial Resistance Australian Enterobacteriaceae Sepsis Outcome Programme annual report, 2014. *Communicable diseases intelligence quarterly report*, 40 (2). E229-E235
6. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13, e1005595.
7. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 2015;5:8365
8. Altschul, S. F., *et al*. Gapped Blast and PSI-Blast: A new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997)
9. NF Alikhan, NK Petty, NL Ben Zakour, SA Beatson (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons, *BMC Genomics*, 12:402. PMID: 21824423
10. Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)*, *27*(7), 1009–1010. doi:10.1093/bioinformatics/btr039
11. Andrew J. Page, Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T.G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, Julian Parkhill, Roary: rapid large-scale prokaryote pan genome analysis, *Bioinformatics*, Volume 31, Issue 22, 15 November 2015, Pages 3691–3693, https://doi.org/10.1093/bioinformatics/btv421
12. Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15, 524.
13. 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 Res. 2016 Jul 8;44(W1):W242-5. doi: 10.1093/nar/gkw290. Epub 2016 Apr 19. PubMed PMID: 27095192; PubMed Central PMCID: PMC4987883.
14. Gordienko, E. N., Kazanov, M. D., & Gelfand, M. S. (2013). Evolution of pan-genomes of *Escherichia coli*, *Shigella spp*., and *Salmonella enterica*. *Journal of bacteriology*, *195*(12), 2786–2792. doi:10.1128/JB.02285-12
15. Snipen, Lars, and David W Ussery. "Standard operating procedure for computing pangenome trees." *Standards in genomic sciences* vol. 2,1 135-41. 28 Jan. 2010, doi:10.4056/sigs.38923
16. Yang, Q. E., Sun, J., Li, L., Deng, H., Liu, B. T., Fang, L. X., … Liu, Y. H. (2015). IncF plasmid diversity in multi-drug resistant *Escherichia coli* strains from animals in China. *Frontiers in microbiology*, 6, 964. doi:10.3389/fmicb.2015.00964
17. Humeniuk, C., Arlet, G., Gautier, V., Grimont, P., Labia, R., & Philippon, A. (2002). Beta-lactamases of *Kluyvera ascorbata*, probable progenitors of some plasmid-encoded CTX-M types. *Antimicrobial agents and chemotherapy*, *46*(9), 3045–3049. doi:10.1128/aac.46.9.3045-3049.2002

18. Poirel L., Kämpfer P., Nordmann P. (2002). Chromosome-encoded Ambler class A β-lactamase of *Kluyvera georgiana*, a probable progenitor of a subgroup of CTX-M extended-spectrum β-lactamases. Antimicrob. Agents Chemother. 46, 4038–404010.1128/AAC.46.12.4038-4040.2002
19. Brooks, L., Kaze, M., & Sistrom, M. (2019). A Curated, Comprehensive Database of Plasmid Sequences. *Microbiology resource announcements*, *8*(1), e01325-18. doi:10.1128/MRA.01325-18
20. Darling AE, Worden P, Chapman TA, Roy Chowdhury P, Charles IG, Djordjevic SP. The genome of Clostridium difficile 5.3. Gut Pathog. 2014 Feb 24;6(1):4. doi: 10.1186/1757-4749-6-4. PubMed PMID: 24565059; PubMed Central PMCID: PMC4234979.

5.6 Figures, tables and supplementary data:

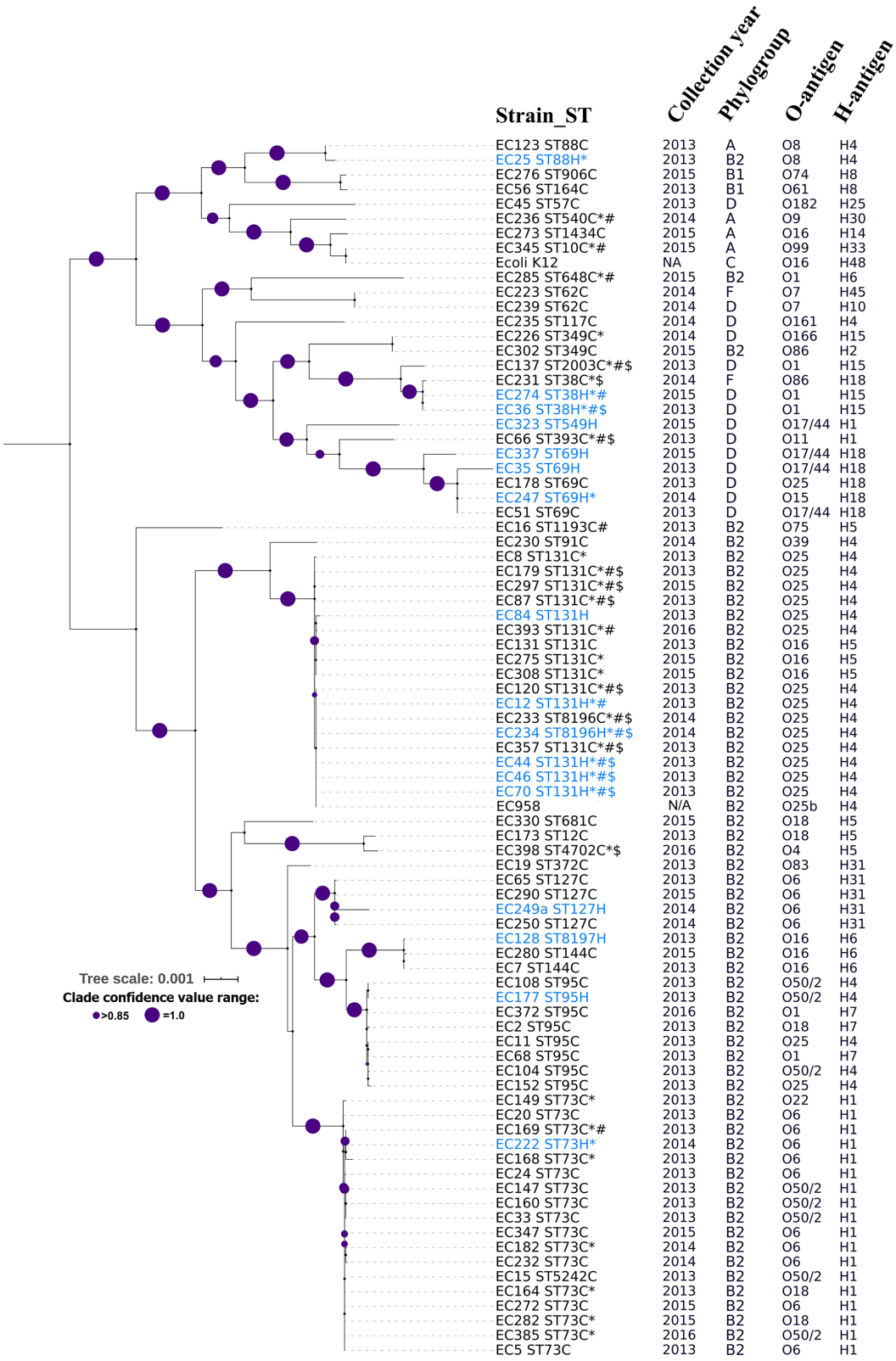
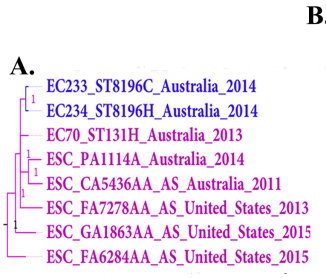


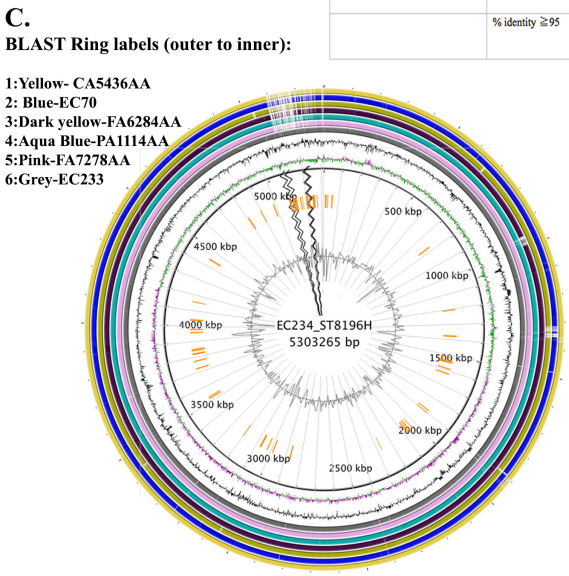
Figure 1:SNP phylogeny of the core genome of sequence types that comprise ST131 clonal cluster. The tree includes ST131 reference genome EC985, and ST131 isolates form our

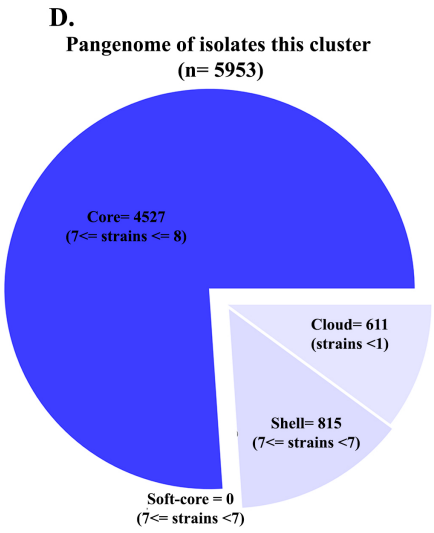
extended collection reported in Hastak et al 2020 (Microbial genomics). Red box indicated the clade which has ST8196 isolates EC233 and EC234.



B.

Genome ID	Serotype/fim H/CH type	Virulence Profile	Resistance Profile	Plasmid Profile
EC70_ST131H	H4:O25/H30/ 40-30	*iha, iss, sat*	*aac(3)-IIa, aac(6')-Ib-cr, aadA5, dfrA17, bla*CTX-M-15, *bla*OXA-1, *sul1, parC* (S80I, E84V)	IncFIA, IncFIB, IncFII
EC233_ST8196C	H4:O25/H30/ 40-30	*iha, iss, sat*	*aac(3)-IIa, dfrA17, bla*CTX-M-15, *bla*DHA-1, *sul1, mphA, qnrB4, parC* (S80I, E84V)	IncFII, IncI1, ColBS512
EC234_ST8196H	H4:O25/H30/ 40-30	*gad, iha, iss, sat*	*aac(3)-IIa, dfrA17, bla*CTX-M-15, *bla*DHA-1, *sul1, mphA, qnrB4, parC (S80I, E84V)*	IncFII, IncI1, ColBS512
ESC_PA1114AA_AS	H4:O25/H30/ 40-30	*gad, iha, iss, sat*	*bla*CTX-M-15, *parC* (S80I, E84V)	IncFIA, IncFIB, IncFII, IncX1
ESC_CA5436AA_AS	H4:O25/H30/ 40-30	*iha, iss, sat*	*aac(3)-IIa, aac(6')-Ib-cr, aadA5, dfrA17, bla*CTX-M-15, *bla*OXA-1, *bla*TEM-1b, *sul1, parC* (S80I, E84V)	IncFII, IncFIB
ESC_FA7278AA_AS	H4:O25/H30/ 40-30	*cnf1, gad, iha, iss, sat, senB*	*aac(3)-IIa, aph(3")-Ib, aac(6')-Ib-cr, aadA5, dfrA17, tet(A), mphA, bla*CTX-M-15, *bla*TEM-1b, *sul1, sul2, parC (S80I,E84V)*	IncFIA, IncFIB, IncFII
ESC_GA1863AA_AS	H4:O25/H30/ 40-30	*astA, iha, iss, sat, senB*	*aac(3)-IIa, aph(3")-Ib, aadA5, dfrA17, tet(A), mphA, bla*CTX-M-15, *bla*OXA-1b, *sul1, sul2, parC (S80I,E84V)*	IncFIA, IncFIB, IncFII
ESC_FA6284AA_AS	H4:O25/H30/ 40-30	*iha, iss, sat*	*aac(3)-IIa, aph(3")-Ib, aadA5, dfrA17, tet(A), mphA, bla*CTX-M-15, *bla*OXA-1b, *sul1, sul2, parC (S80I, E84V)*	IncFIB, IncFII, ColBS512
	% identity ≧95	% identity ≧98, % coverage≧80	% identity ≧98, % coverage≧80	% identity ≧95, % coverage≧80





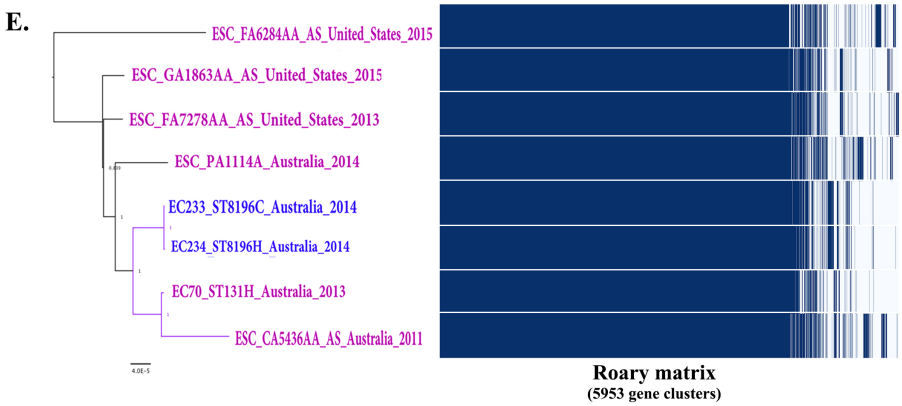


Figure 2:Pangenome analysis of isolates that closely clustered in Mr Bayes analysis (Files S1_Chapter5 and Figure S2). A. The Mr Bayes subclade which consists of closely related ST131 genomes, with year and country of isolation. B. Genotype of isolates generated through the Centre of Genomic Epidemiology portal. C. BLASTn alignment (using 90%

lower and 100% upper identity cutoffs) of all 8 genomes against the partially completed EC234_ST8196 genome co-assembled using short and long reads. Identity droppings below 90% found in dark yellow (FA6284AA), green (GA1863AA) and blue (EC70_ST131) rings D. Graphical representation of peptide clustering representing core and accessory genome of the isolates generated by Roary. E. The pangenome matrix indicating relative presence and absence of peptides clusters representing the core and accessory genes in all 8 genomes and a binary phylogenetic tree based on the presence and absence of accessory genes in the genomes in the left had panel.

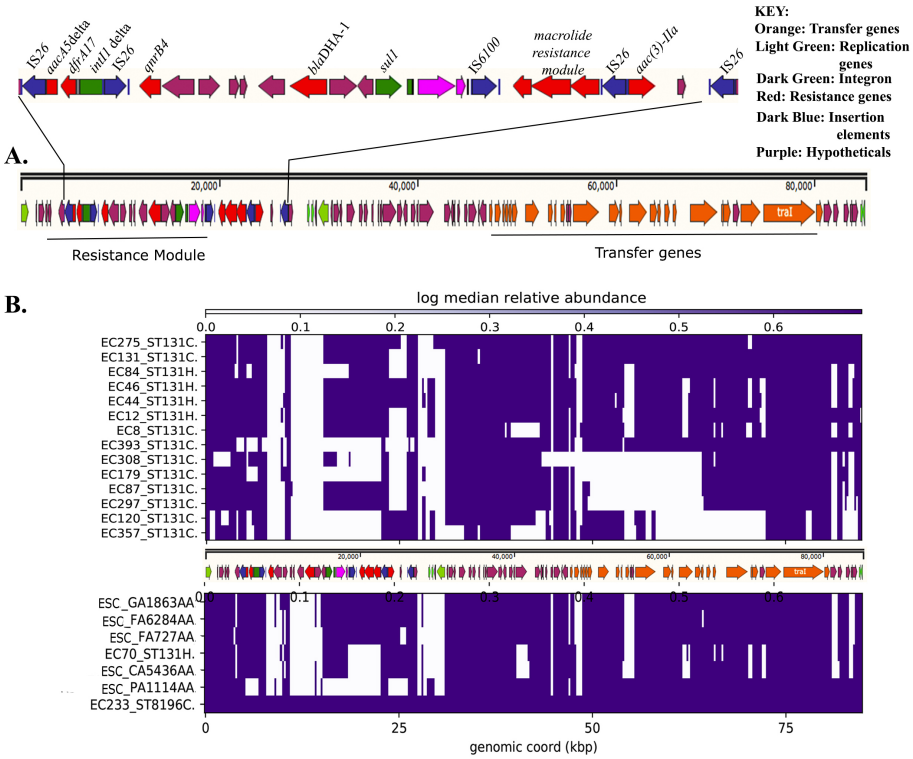


Figure 3: IncFII plasmid analysis. A. Diagrammatic representation of gene content in IncFII plasmid in isolate EC234 and a detailed composition of the resistance loci in the inset above. Arrows representing gene families are colour coded and a key for the colours used is embedded in the figure. B. Mapping of illumina raw reads from ST131 in our collection (upper panel) and EC233_ST8196 and other ST131 isolates identified to be closely related to EC234 from our phylogeny analyses (bottom panel). The purple colour indicates presence of the region or genes and white indicates absence of the region within the analysed plasmid.

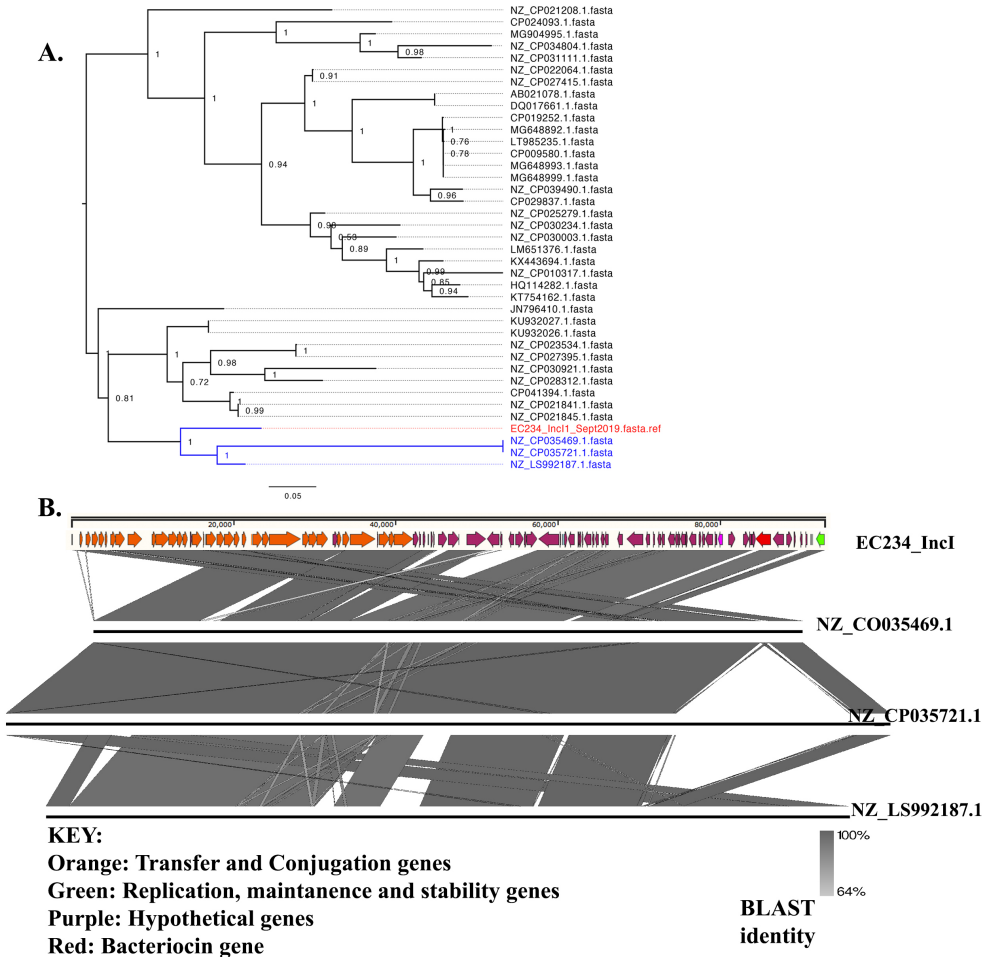


Figure 4: IncI plasmid analysis. A. Phylogeny of complete IncI plasmids available in the plasmid database. B. EasyFig alignment of the three most closely related plasmids identified in isolate EC234.

S1_Chapter5: Meta-data sheet with all cc131 *E. coli* isolates downloaded from Enterobase for Mr Bayes phylogenetic analysis (find attached separate excel sheet file named as S1_Chapter5).

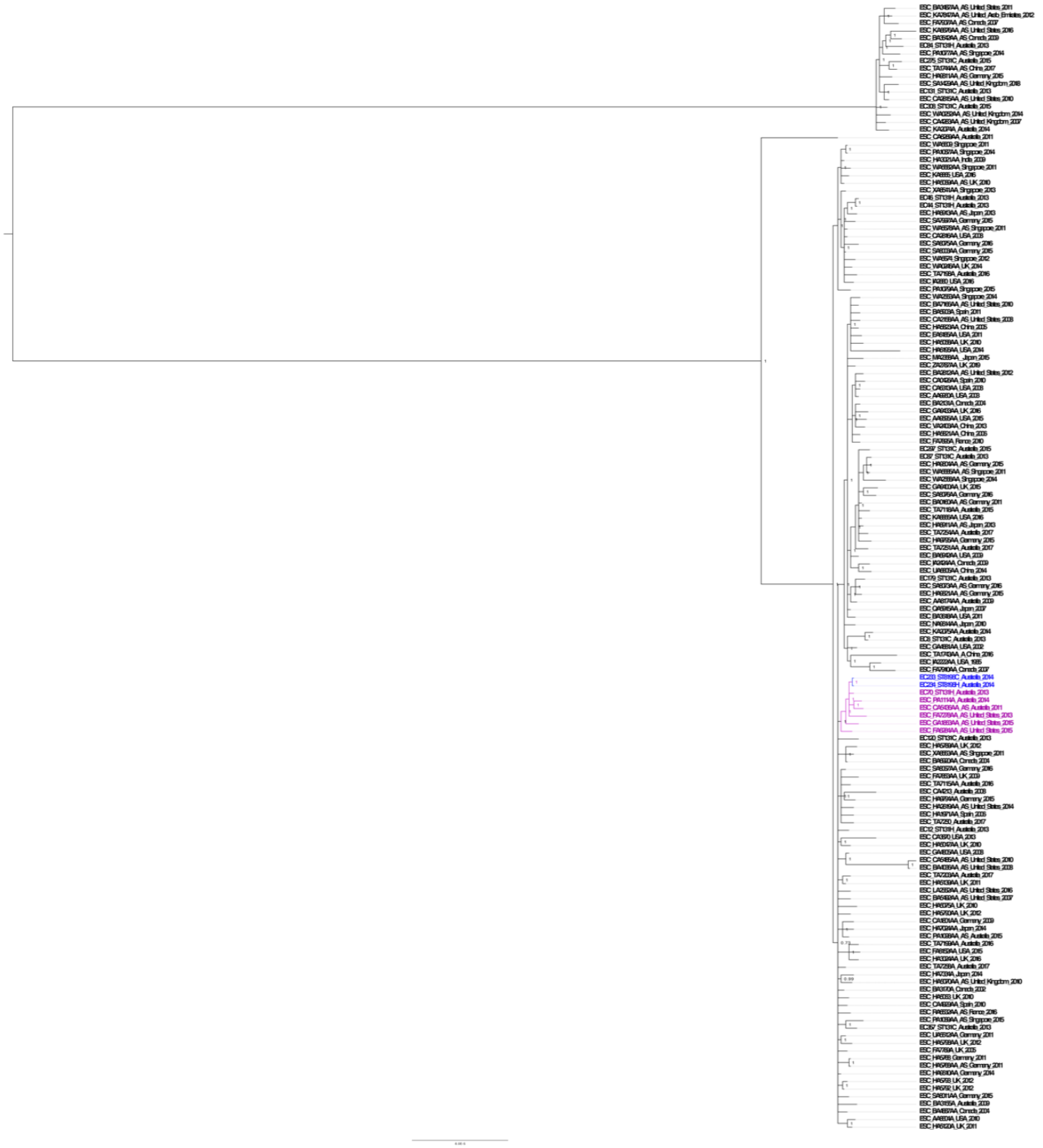


Figure S2: Mr Bayes phylogenetic analysis for determining molecular evolution of cc131 *E. coli* strains.

S3_Chapter5: Island viewer results for the long read sequenced ST8196 strain, EC234_ST8196H (find attached separate excel sheet file named as S3_Chapter5).

S4_Chapter5: Meta-data sheet of presence and absence of genes generated for pan-genome analysis (find attached seperate excel sheet file named as S4_Chapter5)

IncI1 plasmid Accession Number	Host
AB021078.1	Shigella sonnei
CP024093.1	Escherichia coli
CP027395.1	Escherichia coli
CP039490.1	Salmonella enterica subsp. enterica serovar Bareilly str.
JN796410.1	Salmonella enterica subsp. enterica serovar Enteritidis
MG648999.1	Escherichia coli
CP009580.1	Escherichia coli
CP010317.1	Escherichia coli
CP019252.1	Escherichia coli
CP021208.1	Escherichia coli
CP021841.1	Escherichia coli
CP021845.1	Escherichia coli
CP022064.1	Salmonella enterica
CP023534.1	Escherichia coli
CP024993.1	Escherichia coli
CP025279.1	Salmonella enterica
CP027415.1	Salmonella enterica subsp. enterica serovar Typhimurium
CP028312.1	Salmonella enterica subsp. enterica serovar Heidelberg
CP029837.1	Salmonella enterica subsp. enterica serovar Typhimurium
CP030003.1	Salmonella enterica subsp. enterica serovar Brandenburg
CP030234.1	Salmonella enterica
CP030921.1	Escherichia coli
CP031111.1	Escherichia coli
CP034804.1	Escherichia coli
CP035469.1	Escherichia coli
CP035721.1	Escherichia coli
CP038508.1	Escherichia coli
CP041394.1	Escherichia coli
DQ017661.1	Salmonella enterica
HQ114282.1	Escherichia coli
KT754162.1	Shigella dysenteriae
KU932026.1	Escherichia coli
KU932027.1	Escherichia coli
KX443694.1	Escherichia coli
LM651376.1	Escherichia coli
LS992187.1	Escherichia coli
LT985235.1	Escherichia coli
MG648892.1	Escherichia coli
MG648993.1	Escherichia coli
MG904995.1	Escherichia coli

Table S5: IncI1 meta-data sheet of all the downloaded genomes from Enterobase used for IncI1 SNP phylogeny.

Chapter 6: Comparative genomic analysis of ST38 *Escherichia coli* from human, animal and environmental settings reported in Australia and globally

6.3 Introduction:

Pathogenic *Escherichia coli* typically colonize the large intestine in humans and can cause infection in intestinal or extra-intestinal sites. Extra-intestinal pathogenic (ExPEC) *E. coli* are common aetiological agents for urinary tract infections (UTI) and blood-stream infections (BSI). These infections are very difficult to treat due to the rapid spread of antibiotic resistance in ExPEC *E. coli*. The problem is further exacerbated by the overuse of antibiotics in human and animal settings [4] [5]. Although the misuse of antibiotics in human health is a well-known contributor to the rise in antibiotic resistance, the inappropriate administration of antibiotics in food producing animals is now also recognised as a major factor globally [6].

In food producing animals, antibiotics including tetracycline, sulphonamides, macrolides, sulfamethoxazole and penicillin, are often used for growth promotion intensively farmed animals, particularly pigs, poultry and cattle [7]. Moreover, several antibiotics are often batch administered in the feed or water to treat disease outbreaks or for production threatening infections [6]. The overuse of antibiotics in food producing animals results in detectable level of antibiotic residue at slaughter or milking. The contaminant waste effluents from hospitals is another source of antibiotic residues in the environment [8-10] [6]. Notably, these antibiotic residues provide selection pressure that can lead to the uptake of antibiotic resistance and virulence in bacterial populations [8]. This leads to the evolution of multi-drug resistant (MDR) infections resulting in longer hospital stays and increase mortality.

A strain is typically deemed MDR if it is resistant to three or more classes of antibiotics. The rise of carbapenemase-producing and ESBL producing *E. coli* has been widely reported in the recent decade, both within Australia and world-wide. Moreover, there is a dearth of new antibiotics to treat ESBL and carbapenemase-producing Gram-negative bacteria, and effective antibiotic treatment options are dwindling [11].

MDR ExPEC strains are known to harbour the CTX-M gene; the most pandemic ESBL gene is *bla*$_{\text{CTX-M-15}}$, which has been reported in Australia, the United Kingdom, China, India, the United States and Europe. ExPEC strains also harbour a number of ESBL-encoding genes, namely, *bla*$_{\text{OXA-1,-48}}$ and *bla*$_{\text{CTX-M-9,-15,-27}}$. ExPEC strains bearing these genes belonged to phylogenetic group B or D. As ESBL positive isolates are typically MDR they confer resistance against not only beta lactam antibiotics but also a number of non-beta lactam antibiotics, particularly ciprofloxacin, sulfomethoxazole, tobramycin and gentamycin. [12] [13].

In the United States, *E. coli* bearing *bla*$_{\text{CTX-M-14}}$ has been found in a number of different sources such as surface water, fish and vegetables, as well as humans [14]. In Korea, the United Kingdom, Mexico, France and Spain, these ESBL encoding *E. coli* were most frequently identified in urinary tract infections compared to other infections, most of which were nosocomial rather than community acquired infections [15] [16] [1].

Rapid identification of *E. coli* clones associated with ESBL and ARGs, is conducted using multi-locus sequence typing (MLST). This identification defines *E. coli* lineages and provides an insight into the prominent and emerging STs [17]. Globally, the three most prevalent ExPEC clones include ST131, ST38 and ST405 [15]. *E. coli* ST38 clones are increasingly reported in clinical and poultry studies [18] [37], and have also been identified in water bodies, birds and canines. The rapid dissemination of ST38 ExPEC suggests that these clones are on the rise across different ecosystems [19]. Typically, they carry a number of antibiotic resistance genes that encode for resistance against penicillins, cephalosporins, aminoglycosides, trimethoprim and quinolones [20] [21].

Recent clonal outbreaks of *E. coli* ST38 suggest these clones are resistant to carbapenems, the last line of antibiotics for clinical treatment [22]. Globally recognised MDR ST131 *E. coli* are known to be associated with the CTX-M type of ESBLs, however there has been increasing incidence of CTX-M genes within ST38 clones [23].

With the use of whole genome sequencing, it has become easier to screen for genes encoding resistance, virulence factors and the associated mobile genetic elements (MGE). Although a large number of genomic studies have been conducted on ST131 clones, there are limited studies on emerging clones like ST38.

ST38, ST131 and ST405 are major contributors to the distribution of CTX-M producing *E. coli* found in UTIs and BSIs. It is well established that MDR ST131 *E. coli* are typically associated with IS elements such as IS*Ecp1*, IS*26*, IS*CR2*, IS*CR5*, and carry a variety of transposons and plasmid incompatibility groups [1]. ESBL encoding genes have been identified within the chromosome as well as conjugative plasmids in *E. coli*, including IncF and IncI1 plasmids.

Although ST38 MDR *E. coli* has been reported worldwide, there is no studies on whole genome comparisons of these clones in Australia or globally. Here we report the first genomic comparison of three ExPEC MDR ST38 strains, obtained from hospital blood cultures collected from 2013 to 2015 at Concord Repatriation General Hospital, Sydney, Australia. We identified a broad range of AMR genes, virulence factors and MGEs, highlighting the global threat of ST38 *E. coli* as an emerging global MDR clone.

6.4 Materials and methodology:

6.4.1 *E. coli* strains:

E. coli ST38 strains (frozen stocks), EC36_ST38H, EC274_ST38H and EC231_ST38C were collected at Concord Repatriation General Hospital, Sydney from 2013 to 2015 respectively. These strains were collected as blood cultures with no specific source or focus of infection. Two strains were deemed as HO-BSI and one as CO-BSI. The isolates were routinely sub-cultured on Lysogeny Broth (LB) agar plates at 37°C for 18 hours. For genomic DNA extraction, isolates were grown overnight in LB broth on a rotary shaker set at 250 rpm.

6.4.2 Antibiotic resistance phenotyping:

Isolates were screened against a panel of antibiotics: ampicillin (AMP) amoxicillin (AMC), ciprofloxacin (CIP), clofazimine (CFZ), cefotaxime (CTX), ceftriaxone (CRO), ceftazidime (CAZ), cefoxitin (FOX), trimethoprim (TRI), gentamycin (GEN), tobramycin (TOB), norfloxacin (NOR), ticarcillin/clavulanic acid (TCC), piperacillin/tazobactam(TZP) trimethoprim-sulfamethoxazole (SXT), nitrofurantoin (NIT) and amikacin (AMK) and presence of ESBLs (ESBL POS). Antibiotic susceptibility profiles were generated using the Advanced Expert System (AES) software designed for VITEK2 (BioMerieux) system and

interpreted using the European Committee on Antibiotic Susceptibility Testing (EUCAST) breakpoints [34].

6.4.3 DNA extraction, library preparation and whole genome sequencing:

Genomic DNA was extracted from all three strains using 2ml of overnight culture using the ISOLATE II Genomic DNA kit (Bioline, Australia), following the manufacturers protocol. DNA was quantified using the Qubit fluorimeter and dsDNA HS Assay kit (Thermo Fisher Scientific, Australia). Whole genome sequencing libraries were prepared from 2ng of gDNA template with the Illumina Nextera kit. The Illumina HiSeq 2500 v4 sequencer (Illumina, San Diego, CA, USA) in rapid PE150 mode was used for sequencing the genomes. The raw reads generated were assembled into draft genomes using the A5-miSeq pipeline [23]. The short sequences of samples, EC36_ST38H, EC274_ST38H and EC231_ST38C were deposited in GenBank with accession numbers REBF00000000, RDUM00000000 and REBH00000000 respectively.

Long read Pac-Bio microbial multiplex sequencing was performed on two HO-BSI by Ramaciotti Centre for Genomics (UNSW) using a Pacific Biosciences RSII sequencer and HGAP4 assemblies were generated. We characterised antimicrobial resistance regions and complete plasmids from hybrid assemblies generated by Unicycler version 0.4.7 software [24] that utilised the Illumina short raw reads and Pac-bio long reads (https://github.com/rrwick/Unicycler). The long read sequences of EC36_ST38H and EC274_ST38H were deposited in GenBank with bio-sample number SAMN13340720 and SAMN13340721.

6.4.4 Gene identification and MLST:

Antibiotic resistance genes (ARGs) were identified using BLASTn on the UTS high performance cluster, with reference databases ResFinder [25], virulence Finder database (VFDB) [26], and PlasmidFinder [27]. We also screened for the presence of IS elements using ISfinder database. Multi-locus sequence typing (MLST) was conducted using the pubMLST database (https://pubmlst.org/databases/).

6.4.5 Annotations:

Annotations were visualized and managed using SnapGene v3.3.4 (https://www.snapgene.com/). Automated annotations were generated using RASTtk [28] [29].

Insertion sequences were identified and annotated manually using IS finder. Annotations were also conducted manually by BLASTn and using publicly available databases such as ResFinder and the GenBank nucleotide database (https://www.ncbi.nlm.nih.gov/genbank/).

6.4.6 Phylogenetic analysis and figures:

Analysis of core-genome single nucleotide polymorphisms (SNP) was performed using the Harvest suite (Parsnp v1.2, Phipack recombination filter and gingr v1.2) [29]. Core genome alignment was conducted using Harvest ParSNP using the recombination (-x) and iMUM flag (-c) and using the complete genome ESC_SA8299AA_AS isolate as a reference **(Figure 1A)**. Analysis was performed on the IncFII plasmids found in two ST38 strains. Genomic comparisons were conducted using BLASTn and Mauve alignments. Figures were generated using SnapGene v3.3.4 and gingr v1.2 [29].

6.4.7 Comparative Genomics:

Comparative genomics tool used was Mauve version 2.3.1 [30] that utilised tile scaffolds generated by *de novo* A5-miseq assembler to compare against finished EC36_ST38H reference genome. The best alignment generated was selected depending on the highest weight score, indicator of the rearrangements in the genome, and the lowest number of locally collinear blocks (LCBs). Scaffolds that aligned against the finished reference genome were sorted and identified as a subset representing the core genome. The scaffolds that did not align with the finished reference genome were designed as accessory genome.

6.5 Results:

In this study we performed whole genome comparisons of three MDR *E. coli* ST38 strains isolated from Sydney, with 111 previously published ST38 *E. coli* from Australia and globally. Our sequenced cohort included strains isolated from hospital onset infections EC36_ST38H, EC274_ST38H and community onset infection EC231_ST38C.

6.5.1 SNP phylogeny and gene distribution in ST38 *E. coli*:

Total of 114 genomes that included our three sequenced ST38 *E. coli*, EC36_ST38H, EC274_ST38H and EC231_ST38C were used to generate a SNP tree using parSNP software and 67% of their core genome aligned. The SNP tree comprised of smaller subclades comprised of two to ten genomes that grouped together from poultry, chicken, sea-gulls and human origins

and from various countries. The countries include, USA, UK, Germany, Africa, Canada, South America, Norway, Switzerland, China, Thailand, Pakistan and Egypt. Of this, our strains EC36_ST38H, EC274_ST38H and EC231_ST38C belonged to the lower in the SNP tree.

Total number of variant sites used to draw the phylogeny was 37,970. The number of sites in the lower clade were ~7,020 and upper clade were ~ 4,500 and the number of SNP differences in the lower clade were ~5,000 and upper clade were ~3000. This demonstrated that there was high genetic diversity in the lower clade as compared to the upper clade **(Figure 1).**

When we compared our three sequenced ST38 *E. coli*, our hospital onset strains, EC274_ST38H and EC36_ST38H were distantly clustered from the community strain EC231_ST38C **(Table 1)**. The sub-cohort with EC274_ST38H and EC36_ST38H comprised of both hospital and community onset strains from human, sewage and seagulls and from countries like Japan, USA and the UK. EC36_ST38H clustered very closely with ESC_GA8913AA isolated a hospital in the UK in 2015 **(Figure 1)**. These findings, suggested that these clones may be acquired from the community but survive in hospital environments.

Our sequenced strains, EC36_ST38H and EC274_ST38H hospital strains had differences in the core-genome from each other, although both strains were isolated from the same hospital in Sydney and with same source of infection **(Table 1)**. Interestingly, EC274_ST38H hospital strain clustered together with hospital and community strains that were isolated from humans, sewage as well as seagulls from different countries. EC274_ST38H differed by ~1,500 SNPs from isolates, ESC_OA6281AA hospital isolate, ESC_MA2322AA community acquired strain and ESC_YA1551AA hospital strain. This suggests that the ST38 hospital strain was genetically diverse from the other two strains. However, ESC_OA6281AA hospital isolate from urine sample in the UK in 2017 differed by less than 150 SNPs from ESC_MA2322AA isolated from sewage water in Japan in 2015, and MA2322AA differed by less than 35 SNPs from ESC_YA1551AA isolated from a hospital in Switzerland in 2014, indicating these three isolates from community and hospital settings are genetically related in their core genome. This suggests that these strains found in three different countries are conserved may differ in acquisition of accessory genes via lateral gene transfer.

ESC_RA8222AA and ESC_RA8250 isolated in 2014 from humans in the USA had no SNP differences. These two strains from USA and an Australian strain ESC_PA1115AA differed

by less than 10 SNPs, indicating these strains share genetic related in their core genome. Notably, ESC_PA115AA isolated differed by 30 SNPs from ESC_TA2235AA strain isolated from seagull droppings from Australia itself. This demonstrated that the hospital strain and seagull strain are closely related based on their core genome. This may suggest that these sea-gulls maybe vectors for these strains, however to confirm these findings further genomic analysis is conducted later on in this chapter.. When we compared this seagull strain with the other ST38 *E. coli* strains isolated from humans in USA and Switzerland, it differed by ~1,000 SNPs from other strains, indicating genetic diversity between the Australian seagull strain and human strains isolated from other countries.

Our community onset strain, EC231_ST38C formed a separate sub-clade with four other ST38 strains in the SNP tree (**Figure. 1**). Our strain differed by less than 200 SNPs from another MDR hospital strain isolated from an outbreak in Brisbane in 2017 and less than 250 SNPs from a UTI strain isolated in Norway in 2016. This indicates that our community onset strain and to the Australian and Norwegian hospital strains share a conserved core-genome. In order to determine if the ST38 *E. coli* strains from this study are clonal, we need to futher conduct pangenome analysis for these strains.

6.5.2 Antimicrobial resistance and virulence factors in ST38 *E. coli*:

We identified about 12 different ARG's in the global collection of ST38 *E. coli* (S1_Chapter4). From these about 80% harboured more than 3 ARG's (S1_Chapter4). The most common ARGs (50/111) in this collection included; blaTEM-1 that encodes for ampicillin resistance, the co-linked streptomycin resistance gene *strA/B*, sulphonamide resistance gene *sul2*, followed by tetracycline resistance gene *tet(B)*, aminoglycoside encoding genes, *aadA1, aadA2* and *aac(3)-I*. Fluroquinolone encoding gene *floR* that is typically localised in plasmids was identified only in one ST38 *E. coli*. More than 50% of the strains carried ESBL gene, that included CTX-M-9,-15,-25,-27. We screened for presence of clinical class 1 integrase gene *intI*1 gene, and we identified this gene in more than 50% (56/114) of the strains. ARGs that were associated with clinical class 1 integrons included *sul1* (56/114), *aadA* and *dfrA* variants. These genes were predominately found in the sub-clade carrying our two sequenced ST38 strains (**Table.1**). This included the two ST38 strains, EC36_ST38H and EC234_ST38H along with other closely clustered ST38 strains isolated from human source. We also identified upto 10 virulence associated genes and majority (60%) carried *fimH* gene encoding for type 1 fimbriae adhesin, followed by *fyuA* gene associated with UPEC virulence and *iucD* iron acquisition system in

this clade. This demonstrates that the strains within this clade carry similar antimicrobial and virulence genes indicating potential clonal inheritance. However, further bioinformatics analysis is required to determine if these genes are located in the chromosome or on mobile genetic elements as this would provide a better insight to clonal inheritance of these clustered ST38 strains.

Clinically significant sulphonamide resistance is known to disseminate by the acquisition of variants of sulphonamide-resistant dihydropteroate synthases. The *sul1* and *sul2* genes as mentioned before are commonly found in class 1 integrons. Another clinically important sulphonamide resistant gene *sul3* has been recently associated with class 1 integrons. The class 1 integron associated *sul3* was first identified in a Swedish porcine *E. coli* in 2003 and in an Australian commensal ST95 *E. coli* in 2017 (Grape et al, 2003; Reid et al., 2017). The *sul3* structure consists of a putative transposase tnp440, *sul3, orfA/B,* and *mefB* (fragment) and IS*26*. Previous reports have identified the *mefB* fragments of 111bp or 250bp due to IS*26* disrupting the gene. In our study we report a *sul3* module that comprises of *mefB* fragment of 350 bp in three published ST38 *E. coli* isolated from human source in Australia, Thailand and Italy, respectively.

We compared the virulence and resistance profiles with the SNP phylogeny and identified that the poultry strains in the upper clade **(Figure 1)** isolated from Germay, USA and Denmark all carried *fimH, iucD* associated with iron acquisition and *papI* associated with pathogenicity. As these poultry strains carried common virulence genes, this indicated that these ST38 clones found in avian collections persisted and survived in different geographic location potentially due to these specific virulence factors that these strains have acquired. Similarly, we identified a number of isolates such as ESC_RA5159AA and ESC_QA3828 that differed by 20 SNPs and both these strains did not carry any ARGs associated with *intI*1 CRL or IS*26*. However, Australian human strains that formed a close cluster in this clade but varied in their virulence and ARG carriage. For instance, ESC_CB9396AA that closely clustered with ESC_CB9394AA and both carried a number of virulence genes such as *fimH, fyuA*, but the latter did not carry *iss*. Similarly, the ARG patterns varied in the sub-clade of Australian strains, with majority of the strains not carrying any ARGs but carrying clinically important ESBL gene, CTX-M-15 (4/7).

In contrast, the lower clade, a number of closely related strains had similar ARGs and virulence patterns. For example, when we compared ESC_TA2235AA isolated from Australian seagulls with ESC_PA115AA Australian hospital strain, they both carried a number of virulence genes such as *fimH, fyuA, sit, papB, iucD, iutA* and ARGs associated with *intI*1, such as *sul1* and *sul2.* As these both strains were similar in their core genome, as described before, and also carried common ARG and virulence genes, these finding indicate these strains maybe clonal. Notably, ESC_TA2235AA had presence of *sul3* sulphonamide variant, that is discussed further on in the result section.

We then compared the antibiotic resistance, virulence, IS element and plasmid carriage of our sequenced ST38 *E. coli* from BSI with closely related strains obtained from our core-genome SNP analysis as this would provide an insight if these strains have similar or same pattern of acquired genes (Table.2).

As previously mentioned EC231_ST38C community strain was closely related to ESC_CB93977AA Brisbane outbreak strain and ESC_XA7366AA strain from Norway. EC231_ST38C and ESC_CB9397AA had major differences in the acquired resistance gene pool. But the community acquired strains EC231_ST38C and ESC_XA7366AA had same ARG's with presence of clinical class 1 integron intI1 gene and associated ARG's that included *aadA5, dfrA17, mph(A), sul1, sul2, strA/B, chrA*. ESC_CB93977AA did not carry *intI*1 gene but only had the presence of *sul2, strA/B* gene. Both ESC_CB93977AA and EC231_ST38C shared similar virulence genes, that includes iron acquisition gene *irp2*, type 1 fimbral adhesin, *fimH*, serum secretin, *iss*, enzyme glutamate decarboxylase, *gad* and type-1 non-adhesin *nfaE* and *capU*. All three strains had the same seroptype and plasmid replicons and carried the pandemic ESBL gene CTX-M-27.

Similarly, our other hospital strain EC36_ST38H was closely related to ESC_GA8913AA from UK and shared the same resistance profile that included ARGs associated with clinical class 1 integron. They both shared the same serotype, O1:H15, and carried the globally recognised ESBL encoding gene, CTX-M-15. Although the two strains are collected from different countries and from different sources, they share similar acquired genes.

6.5.3 Genomic comparison of closely related ST38 *E. coli*:

Mauve multiple alignment comparisons were performed on the closely related ST38 *E. coli* strains (from SNP analysis) and the scaffolds; the finished EC36_ST38H genome was used as a reference (**Figure S3**).

The common feature between all the genomes except for ESC_GA8913AA was the presence of a scaffold carrying K88 minor fimbrial subunit precursor and insertion elements such as IS*1*, IS*26* and Tn*3* transposon that were identified by BLASTn analysis. In the accessory genome, EC274_ST38H strain had a temperate phage P1 scaffold that harboured the P1 phage machinery; *eae* protein or intimin seen in EPEC and EHEC strains; replication gene RepE and RepL; doc (death on curing) toxin protein. This P1 phage was not identified in our other hospital strain EC36_ST38C or the community strain EC231_ST38C.

BLASTn analysis of the P1 phage indicated high similarity with regions identified in two different scaffolds belonging to two different plasmids; pMH13-051M_5 (accession number: AP018576) and p1_00020 exhibiting 99% sequence identify; 100% and 81% sequence query coverage respectively. The strain with pMH13-051M_5 plasmid was a carbapenem resistant *E. coli* isolated from blood culture in Vietnam in 2013. The p1_000200 was identified in an *E. coli* strain from China. The p1_00200 and pM13-051M_5 along with the four closely related strains contained the phage associated scaffold that were putatively comprising of >90% phage; toxin genes; plasmid machinery and hypothetical gene content. This data suggests that these genes reside on a plasmid.

Except for one of the hospital isolate OA6281AA_AS from urine, all genomes had the presence of a scaffold carrying a number of virulence associated genes, that includes, *papI, iagA* invasion protein; along with MGE, IS*1*. All the genomes carried a scaffold with plasmid conjugative transfer (*tra*) genes however, this scaffold lacked plasmid replicons and any other mobile genetic elements.

Notably, two hospital strains and one community acquired strain, EC274_ST38H, ESC_MA2322AA_AS and OA6281AAA_AS had a scaffold carrying virulence associated genes, *vag* operon, *mcmM* gene; two different toxin-antitoxin system, *ccdB, ccdA* and *PemI* and *PemK*; and notably, conjugative transfer (*tra*) genes associated with IncFII plasmids. BLASTn analysis identified this scaffold in three different plasmids from three different strains isolated from India, Saudi Arabia and Thailand; pV251-a (accession number: LC056447),

pCRE1.1 (accession number: CP034396) and O25b:H4 plasmid (accession number: CP015086) exhibiting 99% sequence identity and sequence coverage of 55%, 94% and 82% respectively. This suggests that these virulence associated plasmids found in these ST38 strains globally are spreading and evolving rapidly which may result in the ST38 clones becoming a dominant or well-established clone over time.

The characterisation of surface antigens by serotyping revealed that the closely related *E. coli* from the sub-cohort with the hospital isolates shared O1:H15 serotype and the other sub-cohort with the community isolate shared O86:H18. The serotype O86:H18 is an extra-intestinal virulence marker and is commonly seen in enteropathogenic strains particularly in adults and children with entero-aggressive *E. coli* infections with diarrhea (Piva et al., 2003). Notably, a surveillance study in Japan revealed O86:H18 has been reported as a predominant serotype in CTX-M-14 carrying ExPEC strain that belonged to ST38. These *E. coli* clones were associated with outbreaks as well as sporadic cases in many hospitals in Japan (Suzuki et al., 2009; Riley et al., 2014).

The phylogroup analysis indicated that first sub-cohort, that included our hospital strains belonged to group D and the second sub-cohort that included our community strain belonged to group F based on the updated Clermont classification. The D and F are reportedly sister phylogroups and differ by the carriage of *arpA* house-keeping gene. This indicates that these closely related ST38 *E. coli* carry similar phylogroups. Notably, majority of virulent extra-intestinal *E. coli* causing UTI are reported to belong to phylogroup B and D. Most of the strains with phylogroup D were not only from UTI patients but also from febrile neutropenia patients. Phage identification was carried out using PHAST tool and we identified the P1 phage only in EC274_ST38H isolate within the two sub-cohorts.

To further explore the accessory genome, plasmid typing conducted using plasmid database (as mentioned in methodology section) identified IncFII-IncFIB in more than 60% (5/8) of the two sub-selected cohort. IncFIA-IncFII plasmids were identified in our hospital strain EC274_ST38C and closely related ESC_MA2322AA from Japan. Small plasmids like ColBS512 and/or Col156 was identified in one of our hospital ST38 *E. coli* and two community strains from Japan and UK.

6.5.4 IncFII and IncI1 plasmids circulating in ST38 *E. coli*:

Unicycler hybrid assemblies resolved the closed plasmids present in the three ST38 *E. coli* strains. From these, hospital strain, EC36_ST38H carried IncF plasmid with AMR cargo and an IncI1 plasmid with clinically important extended spectrum beta-lactamases reported in pandemic MDR *E. coli* clones. EC36_ST38H_IncI1plasmid was an 93000 bp circular sequence, EC36_ST38H_IncFII was a 62,264 bp circular sequence with 76 (18 hypotheticals) and 125 (5 hypotheticals) coding sequences (CDS) respectively.

EC36_ST38H_IncFII plasmid had presence of IncFII conjugative transfer genes, *traX, traI* and *RepA1* and *RepA2* replication regulatory proteins. A number of putative proteins, some of the important ones include; TonB ferric transporter system, FTR1 iron permease and *senB* enterotoxin along with two hypotheticals were flanked by MGEs, IS9*1*Δ and IS*3*Δ. Downstream there was presence of cell division FtsZ, Fe^{2+} ABC transporter and copper suppressor protein, scsD with IS*1Δ* in close proximity. From BLASTn analysis this identical region flanked by MGEs was identified in *E. coli* strains from human urine isolated earlier in 2012, 2015 from Utah, USA (CP035517.1). This indicated that the IncFII plasmid has been circulating and spreading widely and is not only specifically found in Australia. This plasmid may have been acquired horizontally by our ST38 *E. coli* strain from the human urine strains in the USA.

The EC36_ST38H_IncI1 plasmid had presence of conjugative transfer tra operon, trbA/C/D genes and fimbrial pil operon. Also carried *repZ* plasmid replicon and plasmid partition par operon, group II intron reverse transcriptase, *ltrA* and plasmid stability protein *StbB*.

The IncFII and IncI1 plasmids are fully mapped and presented in supplementary data in this chapter **(Figure S2A and Figure S2B)**, along with comparisons to genetically related closed plasmids downloaded from GenBank. We also explored if the two plasmids were present in the ST38 *E. coli* cohort.

Three plasmids, p1E41-1 (CP28484.1), pDA33135-139 (CP029577.1) and AR_0104 plasmid (CP020117.1) were found to be closely related to EC36_ST38H_IncFII plasmid with sequence coverage of 89%, 92% and 90% respectively and percentage identity of 99%. The variation in the gene content that included re-arrangement of mobile genetic elements, particularly IS elements like IS*26*, IS*15*DI and hypotheticals. From the plasmid comparison pDA33135-139

was the most similar to EC36_ST38H_IncFII and they differed by the carriage of putative protein FTR1 iron permease, *senB* endotoxin, two hypotheticals and a copy of IS1. This plasmid was isolated from an ST506 *E. coli* isolated from human (unknown source) in Sweden. We also compared p1E41-1 and AR_0104 plasmid and identified differences in the carriage of a copy of IS*15*DIV and putative genes, such as *yihA*, *YcjY* and two hypotheticals. This indicated that these plasmids have acquired and lost genes and this maybe contributed by the IS elements present in these plasmids. Also as previously known, these plasmids are not only circulating in a specific clone as ST38 *E. coli* but also moving across other different *E. coli* sequence types.

EC36_ST38H_IncFII plasmid harboured a class 1 integron associated complex resistance loci (CRL) carrying IS elements like IS*26* and IS*6100* and ARGs that include *sul1, dfrA17, aadA5, chrA,* and *mph(A)* that encode for resistance to clinically important antimicrobials like sulphonamides, trimethoprim, aminoglycosides, chromate and macrolide phosphotransferase respectively.

The chromate resistance gene *chrA* is close to *tnpR* of Tn*21* at one end and the IRt of the 3'-*CS* at the other end. Downstream to this region is the IS*6100* is in close proximity to the *res* site and this region is usually followed by *mph(A)* region. This CRL has been identified in not only clinical settings but also in animals such as porcine and avian environments. Within the ST38 collection, we identified with CRL (85% sequence coverage and 96% identity) in 78% (151/194) of the strains, suggesting that this CRL is common in ST38 *E. coli* not only in Australia but also world-wide.

Notably we identified Tn*3* and IS*Ecp1* mobilising elements upstream to CTX-M-15. MThe other CRL with CTX-M-15 was associated with ISEcp1 transposable unit and Tn*3* transposon. In this structure IS*Ecp1* mediated transposition leads to the disruption of *tnpA* gene of Tn*3* by IS*Ecp1* and associated CTX-M-15. The IS*Ecp1* belongs to IS*1380* family and is associated with other ESBL genes such as CTX-M-1,-2,-9-25 (Smet et al., 2010). Although CTX-M-15 and IS*Ecp1* along with associated Tn3 transposon was identified in a few isolates (6/114), this region was not present on any IncI1 plasmids.

6.6 Discussion:

ST38 *E. coli* are increasingly reported in ExPEC (extra-intestinal pathogenic *E. coli*), particularly in UPEC (uro-pathogenic *E. coli*) strains that exhibit multi-drug resistance (MDR) and carry virulence associated genes that are responsible for colonisation, pathogenicity and bacterial survival [1]. UPEC strains are known to be found in UTI, that most common causative agents of blood-stream infections (BSI) [2]. These clones are known to be producers of clinically important beta-lactamases, such as ESBLs (extended spectrum beta-lactamases) that includes CTX-M-oxacillin types and carbapenemases ($bla_{\text{NDM-1}}$ and $bla_{\text{OXA-1}}$) [3]. In this study we sequenced two MDR hospital and one MDR community onset ST38 *E. coli* isolated from blood-stream infections carrying beta-lactamases resistant to a broad range of beta-lactam antibiotics. We conducted genomic comparison of these isolates with closely related ST38 *E. coli* that were publicly available in Australia and world-wide.

As per our knowledge there are no genomics studies on ST38 *E. coli* and this study is one of the first to conduct genomic comparison of ST38 *E. coli* isolated in Australia and globally. From our results we identified differences in a number of virulence associated genes as well as IS elements and plasmid associated genes. IncFII/IncFIB plasmids are frequently found in our collection, similar to the literature as these plasmids are reported in a number of commensal and pathogenic multi-drug resistant *E. coli* clones from veterinary and clinical settings (Carattoli et al., 2013). Notably, we also identified a P1 phage associated plasmid in only one of our hospital strain EC274_ST38H that carried antitoxin-toxin genes associated with phages, *phd*, *doc* and and conjugation transfer genes identified in IncF plasmids. Notably, this phage associated plasmid has not been reported before in any ST38 *E. coli* strains.

Overall ST38 *E. coli* are genetically diverse across the globe and this diversity in isolated collected from is seen not only in the human settings but also from animals and environmental settings. Notably, from this SNP analysis we could identify diversity between specific strains from different sources within the global ST38 *E. coli* collection. However, we also identified genetically related *E. coli* that clustered together. For instance, an Australian seagull ST38 strain was closely related to an Australian human ST38 strain indicating that these clones are circulating in both human and animal settings.

After screening for the presence of antibiotic resistance, virulence, ESBL encoding genes and plasmids there were a number of the ST38 *E. coli* strains (Table.1) share similar patterns of gene acquisition which may be due to the movement of antimicrobial resistance regions

circulating within the ST38 *E. coli* population. However, we also identified closely related ST38 *E. coli* that differed in the antimicrobial resistance carriage, indicating that there is diversity in the acquired genes within this collection as well. Notably, the most common IS element in our collection was IS*26* (complete sequence and fragments); found in more than 50% of the collection. These IS elements commonly flanked clinical class 1 associated a number of clinically important ARGs and therefore facilitated the mobilisation of these regions. Hence IS*26* plays an essential role in shaping these clinical class 1 associated structures within this studied population. In this study we have specifically focussed on the three sequenced ST38 *E. coli* strains and compared them to a global collection of 114 ST38 isolates to determine if these *E. coli* from BSI share similar virulence and antimicrobial resistance that are present on a number of MGE. In the future the ST38 *E. coli* strains from our collection can be used to compare with other published ST38 *E. coli* further investigated to determine movement and spread of ARG and virulence genes associated with IS*26* in hospital and community settings in Australia.

Pac-Bio sequencing of our two ST38 *E. coli* was essential for characterising MDR and ESBL associated plasmids and to identify if these plasmids are circulating in ST38 *E. coli* in Australia and globally. Our hospital strain carried an IncFII plasmid hosting a number of ARGs associated with clinical class 1 integrons and this CRL was identified in more than 10 isolates isolated in veterinary and clinical settings in countries like Japan, Norway, Netherlands, USA and Australia.

Our results revealed IncI1 plasmid carrying clinically important CTX-M-15 ESBL gene associated with IS*Ecp1* and Tn3 transposon. CTX-M-15 associated with IS*Ecp1* has been reported in a number of clinical isolates world-wide (Smet et al., 2010). The conjugation transfer genes and other plasmid stability genes were identified in majority of Australian isolates but also in isolates from Italy, USA and Germany from poultry and human sources. Notably, we did not identify CTX-M-15 associated Tn*3* in any of the ST38 *E. coli* in the collection. As per our knowledge there are no Australian reports on IncI plasmids carrying CTX-M-15 in hospital onset ST38 *E. coli*.

Therefore, it is evident that ST38 *E. coli* is an emerging threat in clinical medicine, as these strains carry a plethora of globally dispersed extended spectrum beta-lactamases, antimicrobial resistance genes and virulence genes. There is an urgent need to monitor IS*26* and other mobile

genetic elements in ST38 *E. coli* in the future to understand the molecular evolution and genetic architecture that shapes these strains and influences the acquisition of clinically important antimicrobial resistance genes.

6.7 References:

1. Pitout JD. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. Front Microbiol. 2012 Jan 3(9):1–7.
2. Di Venanzio, G., Flores-Mireles, A.L., Calix, J.J. et al. *Urinary tract colonization is enhanced by a plasmid that regulates uropathogenic Acinetobacter baumannii chromosomal genes.* Nat Commun 10, 2763 (2019) doi:10.1038/s41467-019-10706-y
3. Majed F. Alghoribi, Tarek M. Gibreel, Garry Farnham, Sameera M. Al Johani, Hanan H. Balkhy, Mathew Upton, Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic *Escherichia col*i clones in Riyadh, Saudi Arabia, Journal of Antimicrobial Chemotherapy, Volume 70, Issue 10, October 2015, Pages 2757–2762, https://doi.org/10.1093/jac/dkv188
4. Singer R. S. (2015). Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. Frontiers in microbiology, 6, 28. doi:10.3389/fmicb.2015.00028
5. van der Mee-Marquet, N. L., Blanc, D. S., Gbaguidi-Haore, H., Dos Santos Borges, S., Viboud, Q., Bertrand, X., & Quentin, R. (2015). Marked increase in incidence for bloodstream infections due to *Escherichia coli*, a side effect of previous antibiotic therapy in the elderly. Frontiers in microbiology, 6, 646. doi:10.3389/fmicb.2015.00646
6. Landers, T. F., Cohen, B., Wittum, T. E., & Larson, E. L. (2012). A review of antibiotic use in food animals: perspective, policy, and potential. Public health reports (Washington, D.C. : 1974), 127(1), 4–22. doi:10.1177/003335491212700103
7. Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM et al. Genomic Microbial Epidemiology Is Needed to Comprehend the Global Problem of Antibiotic Resistance and to Improve Pathogen Diagnosis. Front Microbiol 2016;7:843.
8. Djordjevic SP, Stokes HW, Roy Chowdhury P. Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Front Microbiol* 2013;4:86.
9. Huang, X., Yu, L., Chen, X., Zhi, C., Yao, X., Liu, Y., Wu, S., Guo, Z., Yi, L., Zeng, Z., and Liu, J.-H. (2017). High Prevalence of Colistin Resistance and *mcr-1* Gene in *Escherichia coli* Isolated from Food Animals in China. Frontiers in Microbiology 8.
10. Mir, R.A., Weppelmann, T.A., Teng, L., Kirpich, A., Elzo, M.A., Driver, J.D., and Jeong, K.C. (2018). Colonization Dynamics of Cefotaxime Resistant Bacteria in Beef Cattle Raised Without Cephalosporin Antibiotics. Front Microbiol 9, 500.
11. Llor, C., & Bjerrum, L. (2014). Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Therapeutic advances in drug safety, 5(6), 229–241. doi:10.1177/2042098614554919
12. Edward R. Bevan, Annie M. Jones, Peter M. Hawkey, Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype, Journal of Antimicrobial Chemotherapy, Volume 72, Issue 8, August 2017, Pages 2145–2155, https://doi.org/10.1093/jac/dkx146
13. McNulty, C., Lecky, D. M., Xu-McCrae, L., Nakiboneka-Ssenabulya, D., Chung, K. T., Nichols, T., … Hawkey, P. M. (2018). CTX-M ESBL-producing *Enterobacteriaceae*: estimated prevalence in adults in England in 2014. The Journal of antimicrobial chemotherapy, 73(5), 1368–1388. doi:10.1093/jac/dky007
14. Muller A, Stephan R, Nuesch-Inderbinen M. Distribution of virulence factors in ESBL producing *Escherichia coli* isolated from the environment, livestock, food and humans. Sci Total Environ. 2016;541: 667–672. 10.1016/j.scitotenv.2015.09.135

15. Shaik S, Ranjan A, Tiwari SK, Hussain A, Nandanwar N, Kumar N, Jadhav S, Semmler T, Baddam R, Islam MA, Alam M, Wieler LH, Watanabe H, Ahmed N. Comparative Genomic Analysis of Globally Dominant ST131 Clone with Other Epidemiologically Successful Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. MBio. 2017 Oct 24;8(5). doi: 10.1128/mBio.01596-17. PubMed PMID: 29066550; PubMed Central PMCID: PMC5654935.
16. Van der Bij A.K. Peirano G. Pitondo-Silva A. Pitout J.D. (2012) The presence of genes encoding for different virulence factors in clonally related *Escherichia coli* that produce CTX-Ms. Diagn Microbiol Infect Dis 72: 297–302.
17. 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. J Clin Microbiol. 2015 Jan;53(1):160-6. doi: 10.1128/JCM.02562-14. Epub 2014 Oct 29. PubMed PMID: 25355761; PubMed Central PMCID: PMC4290915.
18. Berg ES, Wester AL, Ahrenfeldt J, Mo SS, Slettemeås JS, Steinbakk M, Samuelsen Ø, Grude N, Simonsen GS, Løhr IH, Jørgensen SB, Tofteland S, Lund O, Dahle UR, Sunde M. Norwegian patients and retail chicken meat share cephalosporin-resistant *Escherichia coli* and IncK/blaCMY-2 resistance plasmids. Clin Microbiol Infect. 2017 Jun;23(6):407.e9-407.e15. doi: 10.1016/j.cmi.2016.12.035. Epub 2017 Jan 7. PubMed PMID: 28082191.
19. Hu YY, Cai JC, Zhou HW, Chi D, Zhang XF, Chen WL, Zhang R, Chen GX. Molecular typing of CTX-M-producing *Escherichia coli* isolates from environmental water, swine feces, specimens from healthy humans, and human patients. Appl Environ Microbiol. 2013 Oct;79(19):5988-96. doi: 10.1128/AEM.01740-13. Epub 2013 Jul 26. PubMed PMID: 23892737; PubMed Central PMCID: PMC3811354.

20. Merida-Vieyra, J., De Colsa, A., Calderon Castañeda, Y., Arzate Barbosa, P., & Aquino Andrade, A. (2016). First Report of Group CTX-M-9 Extended Spectrum Beta-Lactamases in *Escherichia coli* Isolates from Pediatric Patients in Mexico. PloS one, 11(12), e0168608. doi:10.1371/journal.pone.0168608
21. 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. J Antimicrob Chemother. 2015 Oct;70(10):2757-62. doi: 10.1093/jac/dkv188. Epub 2015 Jul 16. PubMed PMID: 26183183.
22. Gauthier L, Dortet L, Cotellon G, Creton E, Cuzon G, Ponties V, Bonnin RA, Naas T. Diversity of Carbapenemase-Producing *Escherichia coli* Isolates in France in 2012-2013. Antimicrob Agents Chemother. 2018 Aug;62(8). doi: 10.1128/AAC.00266-18. Print 2018 Aug. PubMed PMID: 29866863; PubMed Central PMCID: PMC6105832.
23. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 2015;31(4):587-589.
24. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13, e1005595.
25. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012;67(11):2640-2644.
26. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V et al. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc Natl Acad Sci U S A 2006;103(15):5977-5982.

27. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.
28. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 2015;5:8365.
29. Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15, 524.
30. Darling AE, Mau B, Perna NT: progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one 2010, 5(6):e11147.
31. Piva, I. C., Pereira, A. L., Ferraz, L. R., Silva, R. S., Vieira, A. C., Blanco, J. E., … Giugliano, L. G. (2003). Virulence markers of enteroaggregative *Escherichia coli* isolated from children and adults with diarrhea in Brasília, Brazil. Journal of clinical microbiology, 41(5), 1827–1832. doi:10.1128/jcm.41.5.1827-1832.2003
32. Kumai, Y., Suzuki, Y., Tanaka, Y., Shima, K., Bhadra, R.K., Yamasaki, S. et al. (2005). Characterization of multidrug-resistance phenotypes and genotypes of *Escherichia coli* strains isolated from swine from an abattoir in Osaka, Japan. Epidemiol Infect 133, 59-70.
33. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. Clin Microbiol Infect. 2014 May;20(5):380-90. doi: 10.1111/1469-0691.12646. Review. PubMed PMID: 24766445.
34. Bell, J.M., Turnidge, J.D., Coombs, G.W., Daley, D.A., Gottlieb, T., Robson, J. and George, N. (2016) Australian Group on Antimicrobial Resistance Australian *Enterobacteriaceae* Sepsis Outcome Programme annual report, 2014. Communicable diseases intelligence quarterly report, 40 (2). E229-E235.
35. Reid CJ, Roy Chowdhury P, Djordjevic SP. Tn*6026* and Tn*6029* are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. Plasmid 2015;80:127-137.
36. Grape, M., Farra, A., Kronvall, G. & Sundstro¨ m, L. (2005). Integrons and gene cassettes in clinical isolates of co-trimoxazole-resistant. Gram-negative bacteria. Clin Microbiol Infect 11, 185–192.
37. Smet, A., Van Nieuwerburgh, F., Vandekerckhove, T. T., Martel, A., Deforce, D., Butaye, P., & Haesebrouck, F. (2010). Complete nucleotide sequence of CTX-M-15-plasmids from clinical *Escherichia coli* isolates: insertional events of transposons and insertion sequences. PloS one, 5(6), e11202. doi:10.1371/journal.pone.0011202
38. Chattaway, M. A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M., … Wain, J. (2014). Evidence of evolving extraintestinal enteroaggregative *Escherichia coli* ST38 clone. Emerging infectious diseases, 20(11), 1935–1937. doi:10.3201/eid2011.131845

6.8 Figures, tables and supplementary data:

Figure 1: SNP phylogeny of core genome of published ST38 *E. coli* genomes and three sequenced genomes from BSI (pink). Phylogeny generated using harvest ParSNP tool and the tree is midpoint rooted with values of the confidence interval (CI) for each node.

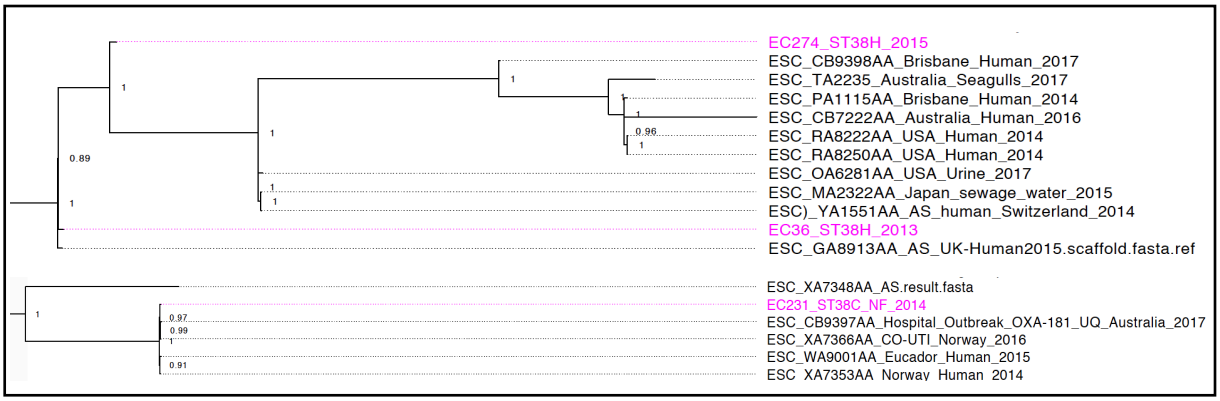


Figure 1B sub-clade 1 (upper clade) with HO isolates and sub-clade 2 (lower clade) with CO isolates. Isolates in pink belong to our sequenced cohort of ST38 *E. coli*. Phylogeny was generated using ParSNP software and the tree is midpoint rooted with values of the confidence interval (CI) for each node.

Sample	Country	Source of infection	SNP differences (from figure. 1)
Subclade 1 with HO isolates:			
EC36_ST38H	Australia	No focus or febrile neutropenia	~150 SNPs (from ESC_GA8913AA)
ESC_GA8913AA	United Kingdom	Human urine	-
EC274_ST38H	Australia	No focus or febrile neutropenia	~550 SNPs (from EC36_ST38H)
EC_MA2322AA	Japan	Sewage water	~1,500SNPs (from EC274_ST38H)
ESC_OA6281AA	United States of America	Human urine	~150 SNPs (from MA2322AA)
Subclade 2 with CO isolates:			
EC231_ST38C	Australia	No focus or febrile neutropenia	-
ESC_CB9397AA	Australia	Human rectal swab	<200 SNPs (from CB9397AA)
ESC_XA7366AA	Norway	Community onset UTI	~250 SNPs (from EC231_ST38C)

Table 1: Comparison of meta-data of closely related ST38 *E. coli* identified from the SNP phylogeny in Figure 1B (two selected sub-cohorts). This table provides information

on the country of sample collection, source of infection and SNP differences determined from the SNP phylogeny.

Sample	Antimicrobial resistance genes	ESBL genes	IS*26*	*intI*1	Virulence associat
Subclade 1 with HO isolates:					
EC36_ST38H	*aac-3-IId, aadA5, dfrA17, mph(A), chrA, sul1, strA/B, sul2, tet(A) TEM-1B*	CTX-M-15	Yes	Yes	*fimH, fyuA, irp2, is papI, iucD, iha, iut senB, hek, pic, pa yeeT*
ESC_GA8913AA	*aac-3-IId, aadA5, dfrA17, mph(A), chrA, sul1, strA/B, sul2, tet(A) TEM-1B*	CTX-M-27	Yes	Yes	*fimH, fyuA, irp2, is kpsM, iucD, iha, i senB, hek, pic,*
EC274_ST38H	*aac-3-IId, catA1, tet(B), TEM-1B*	-	Yes	No	*fimH, fyuA, irp2, is kpsM, papI, iucD, sat, vat, senB, hek,t air, yeeT*
EC_MA2322AA	*aadA5, dfrA17, mph(A), chrA, sul1, strA/B, sul2, tet(A)*	CTX-M-27	Yes	Yes	*fimH, fyuA, irp2, pa iha, iutA, sat, sen, t eilA*
ESC_OA6281AA	*sul1, aadA2, catA1*	CTX-M-15	Yes	Yes	
Subclade 2 with CO isolate:					
EC231_ST38C	*aadA5, dfrA17, mph(A), chrA, sul1, strA/B, sul2, tet(A)*	CTX-M-27	Yes	Yes	*fimH, fyuA, irp2, is vat, papA, air, eat, capU*
ESC_CB9397AA	*Sul2, strA/B*	CTX-M-27	Yes	No	*fimH, fyuA, pic, pa tsh, capU*
ESC_XA7366AA	*aadA5, dfrA17, mph(A), chrA, sul1, strA/B, sul2, tet(A)*	CTX-M-27	Yes	Yes	*fimH, fyuA, yeeT, e*

Table 2: ARG, ESBL, virulence, IS*26* and *intI*1 carriage in closely related ST38 *E. coli* identified from SNP phylogeny in Figure 1B (two selected sub-cohorts).

Sample	Serotype	Phylogroup	Plasmid replicon	Phage associated genes
Subclade 1 with HO isolates:				
EC36_ST38H	O1:H15	D	IncFII IncFIB IncI1	-
ESC_GA8913AA	O1:H15	D	IncFII	-
EC274_ST38H	O1:H15	D	IncFIA IncFII	P1 phage
EC_MA2322AA	O1:H15	D	IncFIA IncFII	-
ESC_OA6281AA	O1:H15	D	IncFIA IncFIB IncII	-
Subclade 2 with CO isolate:				
EC231_ST38C	O86:H18	F	IncFII IncFIB	-
ESC_CB9397AA	O86:H18	F	IncFII IncFIB	-
ESC_XA7366AA	O86:H18	D	IncFII IncFIB	-

Table 3: Serotype, phylogroup, plasmid replicons and phage associated genes in closely related ST38 *E. coli* identified from SNP phylogeny in Figure 1B (two selected sub-cohorts).

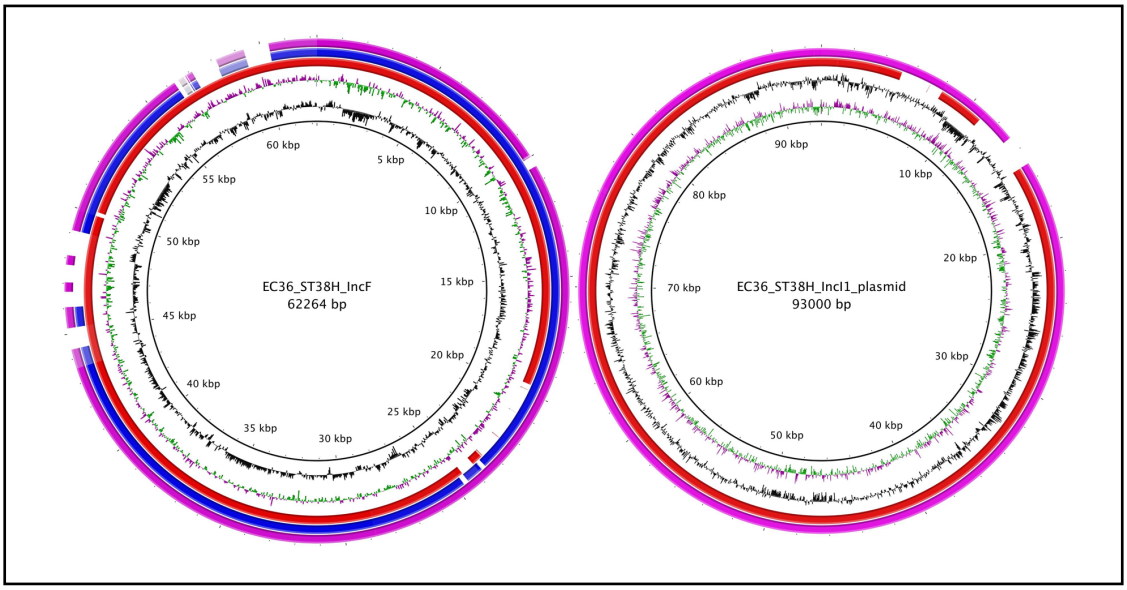


Figure 2: (i) EC36_ST38H_IncF plasmid aligned against pDA33153 (red), p1E41-1 (pink), AR_0104 (blue). (ii) EC36_ST38H_IncI1 plasmid aligned against pSKLX3330 (red), pEK204 (pink). Figure generated using BRIG software and CGView for BLASTn

alignment against reference genome and to graph GC content and skew for the reference.

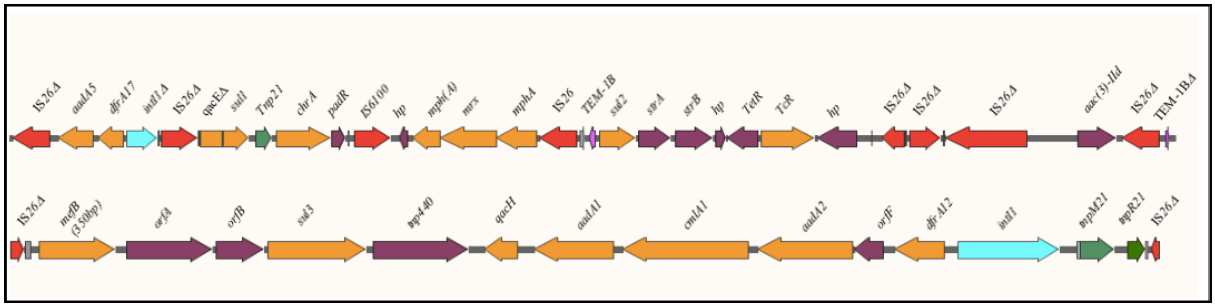


Figure 3A: Sulfonamide resistance genes, *sul1, sul2* and *sul3* and the associated CRL identified in ST38 *E. coli*. Red (IS*26*), orange (ARG), purple (hypotheticals/orfs), grey box (inverted repeats). Figure generated from sequence data in Snapgene software using BLASTn output.

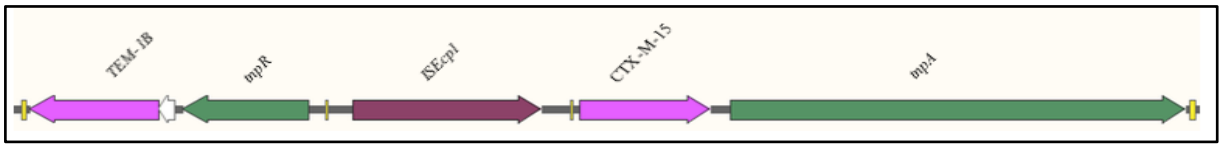


Figure 3B: *bla*$_{\text{CTX-M-15}}$, IS*Ecp1* and Tn*3* identified in ST38 *E. coli*. Red: IS*26*, pink: ESBL gene, green: transposon, yellow: virulence factor, grey/yellow box: inverted repeats. Figure generated from sequence data in Snapgene software using BLASTn output.

Figure S1A. Distribution of clinically important virulence associated genes in ST38 *E. coli*. BLASTn tool used to determine the presence of genes and figure generated

microsoft excel sheet. The purple colour indicates that the gene is present and white indicates gene is absent.

Figure S1B. Distribution of antimicrobial resistance genes in ST38 *E. coli*. BLASTn tool used to determine the presence of genes and figure generated microsoft excel sheet.

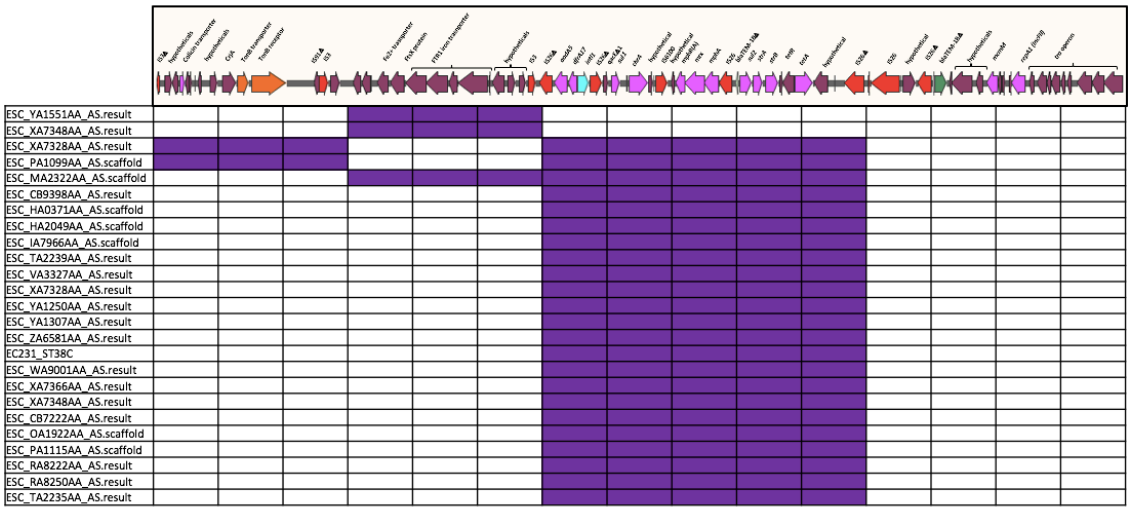


Figure S2A: Plasmid mapping of IncFII (using BLASTn) against our collection of ST38 *E. coli*. Diagrammatic representation of presence or absence of regions in IncFII plasmid generated using BLASTn and plasmid annotations generated in snapgene software.

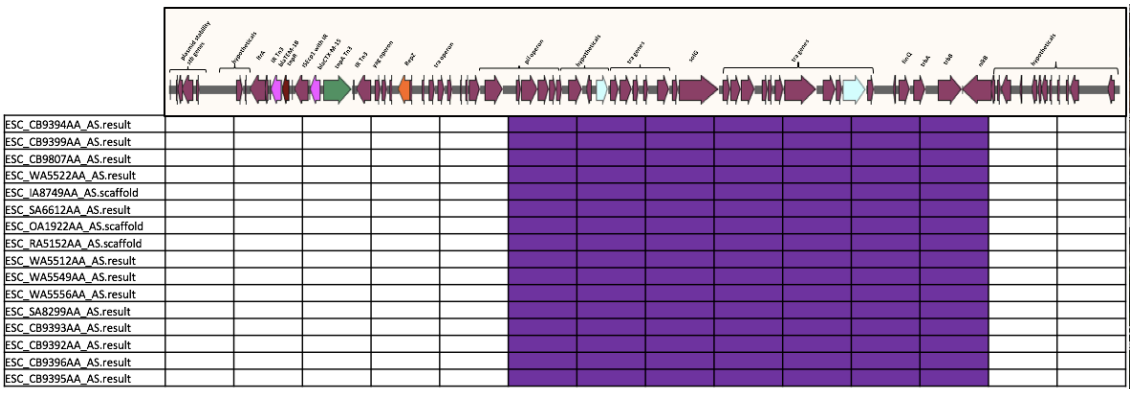


Figure S2B: Plasmid mapping of IncI1 (using BLASTn) against our collection of ST38 *E. coli*. Diagrammatic representation of presence or absence of regions in IncI1 plasmid generated using BLASTn and plasmid annotations generated in snapgene software.

S4_Chapter6: Meta-data of ST38 *E. coli* strains downloaded from Enterobase website find attached seperate excel sheet file named as S4_Chapter6

Chapter 7: General Discussion

In this thesis, each of the Result chapters include a detailed discussion, hence this final chapter is a general discussion on the prevalance of HO and CO isolates, AMR carriage and associated MGE, and virulence carriage and clones circulating in BSI. This section also addresses the clinical significance of this research in Australia and world-wide. The limitations, biases and future challenges will also be outlined.

E. coli is one of the most common *Enterobacteriaceae* that causes hospital and community onset BSI in Australia and in Canada, Iceland, Finland, New Zealand, Sweden, USA, China and Denmark [1]. The rapid spread of antimicrobial resistance, particularly towards antimicrobials that are frequently used to treat *E. coli* infections such as fluroquinolones, cephalosporins and carbapenems, has lead to evolution of MDR *E. coli*. Recent surveillance studies indicate that MDR *E. coli* are on the rise, causing fatal infections such as blood-stream infections. This is a major concern for medical practioners as it reduces treatment options and increases risk of mortality and morbidity [2]. These serious infections are known to carry not only a large cargo of antimicrobial resistance genes but also a plethora of virulence associated genes. However, most of these studies are phenotyping or PCR-based reports that are able to identify only small set of antimicrobial or virulence associated genes depending on targeted primers [3].

Whole genome sequencing (WGS) provides a detailed insight of the carriage of genes, phylogenetic evolution of strains and MGE that may potentially carry ARG and virulence associated genes [7]. This genomic data is critical for investigating strains that are associated with outbreaks in hospitals or within communities. WGS has revolutionised global disease surveillance by rapid outbreak detection and is currently the gold standard in disease epidemiology. With the decreased costs of WGS, it has the capacity to substitute traditional methods to characterise serotypes, virulence, antimicrobial resistance and sequence types in health care laboratories and hospitals. Prior to the work described in this thesis, whole genome sequencing surveillance of *E. coli* BSI had not been described in Australia.

7.1 Prevalance CO and HO *E. coli* BSI:

Chapter 4 examines the collection of 81 *E. coli* isolates from blood-cultures in Concord Repatriate General Hospital, whereas Chapter 5 and 6 focuses on a subset of selected MDR *E. coli* isolates. The blood cultures were dated from 2013 to 2016 with varied sources of infection including intra-abdominal infections, urinary tract infections, biliary tract infections, no focus or febrile neutropenia. Reported BSI are frequently a result of complication of infections in the urinary or gastrointestinal tract. A study by Lefort et al. indicated that death rates were higher in patients with bacteremia that originated from the digestive tract compared that from the the urinary tract. Notably, ExPEC strains are predicted to initate fatal sytemic infections by causing bacteraemia through abdominal translocation in patients with intestinal diseases [4] [5].

BSI are typically classified into community onset or hospital onset infections as associated pathogens, resistance carriage, pathogenictity and clinical outcomes are seen to differ between these two categories [6] [9]. In order to obtain insight into these differences, our cohort were categorised as either CO or HO-BSI. Notably, our cohort had a large number of CO isolates with a MDR phenotype. A recent report by AURA indicate that about 50% of the Australian community were prescribed antimicrobials inappropriately [8]. This may be an important factor in the high carriage of community onset infections as it is well established that the misuse of antimicrobials leads to the evolution of MDR organisms.

An Australian study conducted in Canberra indicated that 68% of BSI were CO as compared to 13% of healthcare-associated BSI [10]. High incidence of CO-BSI is not only seen in Australia and has been reported in a number of other countries. A Canadian study from 2000 to 2006 indicated that the average incidence of BSI was 30.3 infections per 10,000 every year, of which 25.6 per 10,000 was the community onset incidence [1]. Another study in the United States showed that from 1998 to 2007, community associated *E. coli* was found in 59% of cases (Al-Hassan et al., 2009). Notably, a UK hospital study indicated that more than 70% of *E. coli* BSI were non-hospital associated and were identified within 2 days from date of admission. However, it is important to note that these reports varied depending on a number of factors such as antimicrobial treatment history, recency of hospital exposure, geographic location, age of the patient and variation in the categorisation of HO and CO BSI [11].

Although we identified a small proportion of HO isolates in our cohort, these numbers varied depending on the reported source of infection. For instance, in the UK, HO-BSI incidence was

found to be high with UTI patients that have used catheters. More than half of this HO cohort was associated with higher risk of mortality and longer hospital stays in intensive care units (ICU). Another study indicated that gastrointestinal tract infections were frequently identified as the souce of infection in HO-BSI, leading to longer hospital stays [12] [13].

Hospital-associated BSIs were reported as a significant predictor of mortality in multivariable analysis. A study by Diekema et al. found that the mortality rate associated with nosocomial BSI was 34% compared to 14% of CO-BSI [1] [6]. Here we did not have the patient data required to predict if HO or CO BSI were associated with high mortality - this information will be crucial for future research directions. This includes data on the visit of patients to primary care facilities, antibiotic use and treatment regimen, age and gender, geographic location of patient residence, such as postcode or suburb and clinical outcome of the patient [12].

7.2 Epidemiological distribution of *E. coli* BSI:

By conducting WGS, we were able to correlate sequence types, serotypes and phylogroups, and we were able to determine phylogenetic and evolutionary relationships in this cohort. We also identified antimicrobial resistance genes, virulence associated genes and presence of associated mobile elements.

The distribution of phylogenetic groups correlated to the ExPEC clinical population structures indicated that had a largest proportion of of phylogroups were B2 and D followed by A, B1 [94], while strains belonging to group C, E, F were not commonly observed. Notably, the majority of *E. coli* that were members of B2 group belonged to abundant clonal lineages ST131 and ST73, and the well-established STs, ST95 and ST62. However, this phylogroup was not limited to only these STs, as rarely identified STs such as ST681, ST12, ST127, ST114, ST372 also belonged to the B2 group. Two novel STs were identified in this collection: ST8196 that belonged to clonal complex (cc) 131, and ST8197. Both were members of the B2 group. However, the majority (12/27) of rarely identified STs belonged to either phylogroup D or A, with a small proportion belonging to phylogroup B1. The presence of B1 phylogroup in BSI indicates that this group has the ability to cause systemic infections, however this group is more frequently associated with commensal *E. coli* from the intestinal microbiota [14] [15].

Our cohort of 81 sequenced *E. coli*, harboured a number of virulence associated genes required for iron acquisition, pathogenicity, survival and host immunity. According to Fratamico et al.,

the B2 and D groups are associated with high virulence in humans that results in intestinal infections in both healthy and immunocompromised patients. The persistance and survival of *E. coli* strains in BSIs is dependent on serum resistance [15] [16]. Virulence factors associated with serum survival was found in more than half of our collection, reflecting recent literature [17] [15]. Capsule related virulence factors such as *KpsMTII* were found in the majority of this cohort, encoding K1 and K5 capsules. These capsules are polysaccharide antigens that induce host immune responses in BSI. K5 capsules also enable resistance to the human innate and adaptive immune system [18].

Iron uptake is an essential mechanism for catalyzing several indispensable enzymatic reactions. Bacterial siderophores have a high affinity for iron, which allows bacteria to outcompete the host in iron uptake. We identified a number of genes responsble for iron acquisition that included *fyuA, iroN, irp2* and *iucD* in 60% of our collection. Moreover, *fyuA* is a clinically important virulence gene responsible for the invasion of bacteria from urinary tract to the blood-stream [19] [21].

Genes related to adhesins such as type 1 fimbia (*fimH*), P fimbrae (*PapC* and *papG*) were identified in nearly 75% of the cohort sourced from UTI. *FimH* is an important virulence factor found in uropathogenic *E. coli* and UTI bacteraemia . *PapC* is an adhesin that is commonly found in UTIs. Genes associated with toxins and hemolysin production such as *cnf1, hylA, cvaC* and *ibeA* were identified in less than 30% of the cohort [15]. These genes were mainly identified in ST73 *E. coli*, which is one of the most abundant lineage in our cohort. Notably, ST73 *E. coli* harboured highest number of virulence genes compared to any other *E. coli* clones. They carried an average of 7-8 virulence genes that were associated with serum survival, iron acquisition and pathogenicity. ST73 *E. coli* strains may therefore be the dominant lineage in this cohort due to the presence of these virulence genes that play a role in the persistence and survival of these infections.

7.3 Epidemiology of antimicrobial resistance and mobile genetic elements circulating in *E. coli* BSI population:

The data generated from Chapter 4 provided essential information on AMR and clinically important ESBL cargo in *E. coli* BSI in Australia. The genotyping data showed the presence of clinical class 1 integrons (47/81) with known gene cassettes for aminoglycosides (*aadA*), trimethoprim (*dfrA*), macrolide (*mphA*) and chloramphenicol (*cmlA*) resistance. The presence

of sulphonamide resistance genes, *sul1, sul2* and *sul3* were linked to the class 1 integron structure. Tetracycline (*tet*) and penicllin resistance (*blaTEM*) gene variants were frequently identified in our cohort. Notably, ten different complex resistance loci (CRL) were identified carrying clinical class 1 integrase genes. We showed that *sul1* derivatives of clinical class 1 integrons were more common than *sul2* and *sul3*, as also reflected in the literature.

The most prevalant structures carrying ARGs were associated with Tn*21*/Tn*1721* transposon backbones. Notably, these structures also had IS*26* elements flanking them, indicating that IS26 is a driver for the dissemination of ARGs in these BSIs. We also identified IS26 in a number of CRL that carried *sul2, strA/B*, bla_{TEM} and *aphA1* - this structure was associated with the *Tn6026* transposon. Many of the structures seen in our cohort were flanked by IS*26* fragments. The fragments of IS*26* flanking the CRL typically occurred as either the scaffolds ended at these ends or the full copy of IS26 was not present. We identified the *sul3* integron in ST95 *E. coli;* this structure associated with clinical class 1 integron and had a *mefB* remnant of 776bp. This structure carries *dfrA12, aadA* variants, *cmlA1, orfF, orfA/B, qacH* followed by *tnp440*. This genetic signature has been previously identified in another ST95 *E. coli* from an healthy human [22] and in food producing animals in Australia [23], suggesting that this CRL is moving across human and animal settings within the country.

We detected ESBLs that included $bla_{CTX\text{-}M}$ and bla_{OXA} variants in our collection, along with fluroquinolone resistance genes (*qnr* and *floR*). The presence of these ESBL was a sign that antimicrobial stewardship was not well established. This is because fluroquinolones are commonly administered for UTI patients and the misuse of these antimicrobials can lead to selective pressures and consequently the acquisition of ESBL variants, which confers resistance to a number of clinically important antimicrobials. The commonly encountered ESBL gene $bla_{CTX\text{-}M\text{-}15}$ was not limited to the globally disseminated MDR ST131 *E. coli* but also indentified in rare/novel STs such as ST8196, ST38 and ST393.

The most common plasmid circulating in this collection was IncFII-IncFIB, commonly seen in *Enterobacteriaceae* obtained from animals, human and environmental settings. Within the globally disseminated ST131 strains, we identified a ColV-like plasmid that was also seen in 5 ST95 strains, 1 each of ST69, ST88 and ST549. All carried clinically important virulence and colicin resistance genes that are essential for survival of *E. coli* strains in BSI.

7.4 Novel ST8196 *E. coli* identified in BSI:

In Chapter 6, we identified novel ST8196 *E. coli* that clustered phylogenetically with ST131 *E. coli*. This was an important finding as ST8196 could potentially be a new clone circulating within *E. coli* BSIs that has either not been reported before, or has emerged recently. The two ST8196 *E. coli* differed only by 2 SNPs, and both were hospital onset isolates, suggesting that this new ST is circulating in the hospital. In order to determine how closely related the ST8196 strains were to clonal complex (cc) 131 strains in the core genome, we generated a SNP phylogeny in Chapter 6. For this tree we used our ST8196 and ST131 strains, with representatives of each ST belonging to cc131 globally. The tree showed that ST8196 strain EC234_ST131H was most closely related to our ST131 strain, EC70_ST131.

We then wanted to examine if these two strains share the same ancestory or if ST8196 strains had evolved from other cc131 strains identified globally. We applied Mr. Bayes software based on the General Time Reversal (GTR) model and generated a SNP phylogeny using 100 trillion bootstrapping iterations. This tree revealed that our two ST8196 strains, EC234_ST131H and EC233_ST131H shared ancestory with EC70_ST131C. This clearly demonstrates that these strains are circulating in both community and hospital settings.

7.5 Multi drug resistant ST38 *E. coli* an emerging and evolving pathogen:

ST38 *E. coli* are known to carry clinically important AMR cargo and have been increasingly reported in Germany, France, Lebanon and the UK [20]. However, there are no studies on genomic comparisons of this *E. coli* lineage. In our collection we identified three ST38 *E. coli*, of which two were HO and one CO isolate (Chapter 4). In Chapter 5 we conducted genomic comparisions of ST38 *E. coli* from Australia and globally to examine ARG and virulence carriage. We then zoomed into the closely related ST38 strains to determine the similarity of core or acquired genomes, and to identify any potential MGE carrying ARG and other virulence genes.

We conducted this analysis by first generating a SNP tree from 114 ST38 *E. coli* isolates using the parSNP pipeline. This tree revealed a large diversity within these ST38 *E. coli*. Our hospital isolates closely clustered with isolates from other countries such as Norway and Japan.

Notably, we identified ESBL-like *bla*$_{\text{CTX-M-15,-9,-27}}$ in a number of the closely related ST38 *E. coli*. One of our hospital strains carried *bla*$_{\text{CTX-M-15}}$ on an IncI1 plasmid, which has not been reported in IncI1 plasmids in any hospital onset ST38 *E. coli* in Australia or globally. This suggests that this CTX-M-15 gene may have been recently acquired by IncI1 plasmid in a ST38 *E. coli* strain locally in an Australian hospital.

7.6 Summary and limitations:

Genomic epidemiology provides deeper insights for analysing biological organisms via computational tools and biostatistical models to address crucial questions on their pathogenesis, evolution and survival. In this pilot project examining *E. coli* obtained from human blood-stream infections, we conducted WGS on 81 solates, and used this data to determine serotypes, clonality, virulence and antimicrobial resistance genes, plasmids, integron structures and correlated this to associated phenotypic metadata. We sequenced a number of clinically important STs that included MDR-associated ST131 and high virulence factor-hosting ST73 *E. coli* along. We also found emerging MDR STs such as ST117, ST1193 and ST38 that have been reported recently in a number of clinical and avian isolates, as well as novel STs such as ST8196 and ST8197 that have not been previously reported [24-28].

7.6.1 HO and CO classification:

HO and CO classifications are an arbitrary definition, which creates challenges. It is difficult to clearly determine when or where the actual acquisition of infection occurs. In addition, interobserver variation is a source of bias (Mayer et al., 2012). Moreover, the health care delivery system has changed drastically over a short period of time. There is higher levels of complex medical care such as hemodialysis and outpatient antimicrobial treatments delivered in the community itself. Thus, community based patients may acquire hospital associated infections - this blurs the distinction between the two categories [11].

7.6.2 Blood cultures and sample size:

Blood cultures are mandatory to determine BSI diagnosis. Blood cultures typically contain broth media and blood, inoculated under sterile conditions in order to minimise contaminants, and to ensure that the cultured bacteria is associated with BSI. However there is no gold standard for defining a “positive” blood culture. The determinants include timing of the positive culture, presence of other contaminants, and transient bacteremia [29]. If stringent techniques for determining true positives are not conducted, there are chances of “false

positives", where isolates that are not associated with BSI are reported as one, or "false negatives", where a clinically important isolate is missed. Our sample size is relatively small (81 isolates), and may be a narrow representation of the circulating *E. coli* population associated with BSI. This sampe size limitation may affect our analysis and findings [1].

Going forward, it will be essential to increase the sample size to enable larger epidemiological studies and to strengthen evidence for the ongoing development of antimicrobial resistance. As the 'One Health' approach focuses on the interconnections between human, animal, water bodies and environment settings, it is critically important to determine the patterns of AMR carriage and dissemination across these different settings. By identifying these interconnections, we can improve antimicrobial stewardship in human and animal settings to help reduce the spread of AMR infections. It is accepted that wide-spread, inappropriate use of antimicrobials globally has resulted in the rapid evolution of MDR strains. This selection pressure of inappropriately used antibiotics has resulted in exacerbation of infections leading to higher mortality and morbidity globally. Thus there is an urgent need for further rigorous genomic analysis, in the vein of the work presented here, to comprehensively determine the evolution and persistance of bacterial MDR and virulence as an essential component of global AMR reduction strategies.

References:

1. Laupland KB, Gregson DB, Church DL, Ross T, Pitout JD. Incidence, risk factors and outcomes of Escherichia coli bloodstream infections in a large Canadian region. *Clin Microbiol Infect*. 2008 Nov;14(11):1041-7. doi: 10.1111/j.1469-0691.2008.02089.x. PubMed PMID: 19040476.
2. Ibrahim ME, Bilal NE, Hamid ME. Increased multi-drug resistant Escherichia coli from hospitals in Khartoum state, Sudan. Afr Health Sci. 2012 Sep;12(3):368-75. doi: 10.4314/ahs.v12i3.19. PubMed PMID: 23382754; PubMed Central PMCID: PMC3557680.
3. Su, M., Satola, S. W., & Read, T. D. (2019). Genome-Based Prediction of Bacterial Antibiotic Resistance. Journal of clinical microbiology, 57(3), e01405-18. doi:10.1128/JCM.01405-18
4. Lefort A, Panhard X, Clermont O, Woerther PL, Branger C, Mentre F, Fantin B, Wolff M, Denamur E. Host factors and portal of entry outweigh bacterial determinants to predict the severity of Escherichia coli bacteremia. J Clin Microbiol 2011; 49:777-83; PMID:21177892; http://dx.doi.org/10.1128/JCM.01902-10
5. Daga, A. P., Koga, V. L., Soncini, J. G. M., de Matos, C. M., Perugini, M. R. E., Pelisson, M., Vespero, E. C. (2019). Escherichia coli Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. Frontiers in Cellular and Infection Microbiology, 9, 191.
6. Hoenigl M, Wagner J, Raggam RB, Prueller F, Prattes J, et al. (2014) Characteristics of Hospital-Acquired and Community-Onset Blood Stream Infections, South-East Austria. PLOS ONE 9(8): e104702. https://doi.org/10.1371/journal.pone.0104702
7. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol* Rev. 2017 Oct;30(4):1015-1063. doi: 10.1128/CMR.00016-17. Review. PubMed PMID: 28855266; PubMed Central PMCID: PMC5608882.
8. Australian Commission on Safety and Quality in Health Care. AURA 2016: first Australian report on antimicrobial use and resistance in human health. Sydney: ACSQHC; 2016
9. Garner JS, Jarvis WR, Emori TG, Horan TC, Hughes JM. CDC definitions for nosocomial infections, 1988. Am J Infect Control. 1988 Jun;16(3):128-40. doi: 10.1016/0196-6553(88)90053-3. PubMed PMID: 2841893.
10. KJ Kennedy, JL Roberts, PJ Collignon *Escherichia coli bacteraemia in Canberra: incidence and clinical features Med J* Aust, 188 (2008), pp. 209-213
11. Laupland K. B. (2013). Incidence of bloodstream infection: a review of population-based studies. *Clin. Microbiol. Infect.* 19, 492–500. 10.1111/1469-0691.12144
12. Vihta, Karina-Doris & Stoesser, Nicole & Llewelyn, Martin & Quan, T & Davies, Timothy & Fawcett, Nicola & Dunn, Laura & Jeffery, Katie & Butler, Chris & Hayward, Gail & Andersson, Monique & Morgan, Marcus & Oakley, Sarah & Mason, Amy & Hopkins, Susan & Wyllie, David & Crook, Derrick & Wilcox, Mark & Johnson, Alan & Walker, A. (2018). Trends over time in Escherichia coli bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: a study of electronic health records. The Lancet Infectious Diseases. 18. 10.1016/S1473-3099(18)30353-0.

13. Guidance on the definition of healthcare associated Gram-negative bloodstream infections. NHS Improvement and Public Heath England, July, 2017
14. Skjøt-Rasmussen, Line & Ejrnæs, Karen & Lundgren, Bettina & Hammerum, Anette & Frimodt-Møller, Niels. (2012). Virulence factors and phylogenetic grouping of Escherichia coli isolates from patients with bacteraemia of urinary tract origin relate to sex and hospital- vs. Community-acquired origin. International journal of medical microbiology : IJMM. 302. 129-34. 10.1016/j.ijmm.2012.03.002.
15. Daga AP, Koga VL, Soncini JGM, de Matos CM, Perugini MRE, Pelisson M, Kobayashi RKT, Vespero EC. Escherichia coli Bloodstream Infections in Patients at a University Hospital: Virulence Factors and Clinical Characteristics. Front Cell Infect Microbiol. 2019;9:191. doi: 10.3389/fcimb.2019.00191. eCollection 2019. PubMed PMID: 31245301; PubMed Central PMCID: PMC6563721.
16. Fratamico P. M., DebRoy C., Liu Y., Needleman D. S., Baranzoni G. M., Feng P. (2016). Advances in molecular serotyping and subtyping of Escherichia coli. Front Microbiol. 3:644 10.3389/fmicb.2016.00644
17. Koga V. L., Tomazetto G., Cyoia P. S., Neves M. S., Vidotto M. C., Nakazato G., et al. . (2014). Molecular screening of virulence genes in extraintestinal pathogenic Escherichia coli isolated from human blood culture in Brazil. Biomed. Res. Int. 2014:465054. 10.1155/2014/465054
18. Blundell C. D., Roberts I. S., Sheehan J. K., Almond A. (2009). Investigating the molecular basis for the virulence of Escherichia coli K5 by nuclear magnetic resonance analysis of the capsule polysaccharide. J. Mol. Microbiol. Biotechnol. 17, 71–82. 10.1159/000215933
19. Johnson JR, Stell AL. Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise. J Infect Dis. 2000 Jan;181(1):261-72. doi: 10.1086/315217. PubMed PMID: 10608775.
20. Chattaway, M. A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M., … Wain, J. (2014). Evidence of evolving extraintestinal enteroaggregative Escherichia coli ST38 clone. Emerging infectious diseases, 20(11), 1935–1937. doi:10.3201/eid2011.131845
21. Ananias M, Yano T. Serogroups and virulence genotypes of Escherichia coli isolated from patients with sepsis. Braz J Med Biol Res. 2008 Oct;41(10):877-83. doi: 10.1590/s0100-879x2008001000008. PubMed PMID: 19030710.
22. Moran RA, Holt KE, Hall RM. pCERC3 from a commensal ST95 Escherichia coli: A ColV virulence-multiresistance plasmid carrying a sul3-associated class 1 integron. Plasmid. 2016 Mar-May;84-85:11-9. doi: 10.1016/j.plasmid.2016.02.002. Epub 2016 Feb 18. PubMed PMID: 26855083.
23. Reid, C. J., Wyrsch, E. R., Roy Chowdhury, P., Zingali, T., Liu, M., Darling, A. E., … Djordjevic, S. P. (2017). Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. Microbial genomics, 3(12), e000143. doi:10.1099/mgen.0.000143
24. Veronika L Tchesnokova, Elena Rechkina, Lydia Larson, Kendra Ferrier, Jamie Lee Weaver, David W Schroeder, Rosemary She, Susan M Butler-Wu, Maria E Aguero-Rosenfeld, Danielle Zerr, Ferric C Fang, James Ralston, Kim Riddell, Delia Scholes, Scott Weissman, Kaveri Parker, Brad Spellberg, James R Johnson, Evgeni V Sokurenko, Rapid and Extensive Expansion in the United States of a New Multidrug-resistant Escherichia coli Clonal Group, Sequence Type 1193, Clinical Infectious Diseases, Volume 68, Issue 2, 15 January 2019, Pages 334–337

25. 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. J Clin Microbiol. 2015 Jan;53(1):160-6. doi: 10.1128/JCM.02562-14. Epub 2014 Oct 29. PubMed PMID: 25355761; PubMed Central PMCID: PMC4290915.
26. Chattaway, M. A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M., … Wain, J. (2014). Evidence of evolving extraintestinal enteroaggregative Escherichia coli ST38 clone. Emerging infectious diseases, 20(11), 1935–1937. doi:10.3201/eid2011.131845
27. Johnson JR, Johnston BD, Porter SB, Clabots C, Bender TL, Thuras P, Trott DJ, Cobbold R, Mollinger J, Ferrieri P, Drawz S, Banerjee R. Rapid Emergence, Subsidence, and Molecular Detection of Escherichia coli Sequence Type 1193-fimH64, a New Disseminated Multidrug-Resistant Commensal and Extraintestinal Pathogen. J Clin Microbiol. 2019 May;57(5). doi: 10.1128/JCM.01664-18. Print 2019 May. PubMed PMID: 30787145; PubMed Central PMCID: PMC6498021.
28. Majed F. Alghoribi, Tarek M. Gibreel, Garry Farnham, Sameera M. Al Johani, Hanan H. Balkhy, Mathew Upton, Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic Escherichia coli clones in Riyadh, Saudi Arabia, Journal of Antimicrobial Chemotherapy, Volume 70, Issue 10, October 2015, Pages 2757–2762, https://doi.org/10.1093/jac/dkv188
29. Hall, K. K., & Lyman, J. A. (2006). Updated review of blood culture contamination. Clinical microbiology reviews, 19(4), 788–802. doi:10.1128/CMR.00062