Genomic Characterisation of Extra-intestinal Pathogenic *Escherichia coli* from Human Infection

A thesis submitted for the degree of

Doctor of Philosophy

by

Jessica McKinnon

Bachelor of Biotechnology (Honours)

Certificate of Authorship

I declare that this thesis by compilation is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution.

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

Production Note:
Signature removed prior to publication.

Jessica McKinnon

May 2020

Peer Reviewed Publications

Published with First Authorship (* authors contributed equally)

Darling AE*, **McKinnon J***, Worden P, Santos J, Charles IG, Roy Chowdhury P and Djordjevic SP. (2014). A draft genome of *Escherichia coli* sequence type 127 strain 2009-46. *Gut Pathogens* 6(32).

McKinnon J*, Roy Chowdhury P* and Djordjevic SP. (2018). Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis. *International Journal of Antimicrobial Agents* 52(3):430-435.

Roy Chowdhury P*, **McKinnon J***, Liu M and Djorodjevic SP. (2018). Multidrug resistant uropathogenic Escherichia coli ST405 with a novel, composite IS*26* transposon in a unique chromosomal location. *Frontiers* 8(9):3212.

McKinnon J, Roy Chowdhury P and Djordjevic SP. (2020) Molecular analysis of an IncF ColV-like plasmid lineage that carries a complex resistance locus with a trackable genetic signature. *Microb Drug Resist*.

Bogema DR*, **McKinnon J***, Liu M, Hitchick N, Miller N, Venturini C, Iredell J, Darling AE, Roy Chowdhury P, Djordjevic SP. (2020). 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* 6(1).

In Preparation

McKinnon J, Roy Chowdhury P and Djordjevic SP. (2019). Whole genome sequencing of 83 class 1 integron positive clinical isolates from urinary tract infection.

Co-Authorships

Melvold JA, Wyrsch ER, **McKinnon J**, Roy Chowdhury P, Charles IG and Djordjevic, SP. (2017). Identification of a novel qnrA allele, qnrA8, in environmental Shewanella algae. *Journal of Antimicrobial Chemotherapy* 72, 2949-2952.

Roy Chowdhury P, **McKinnon J**, Wyrsch E, Hammond JM, Charles IG and Djordjevic SP. (2014). Genomic interplay in bacterial communities: implications for growth promoting practices in animal husbandry. *Frontiers in Microbiology* 5.

Conference Presentations

- Molecular Dissection of Drug Resistance Regions in Clinical *Escherichia coli* Isolates from a Sydney Hospital

  New Horizons: Research and Education for Optimal Health
  Wednesday 20th November, 2013, Poster 3

- Comparative Genomics of Sequential *Escherichia coli* Strains 2009-49 and 2009-52: A Micro-Evolutionary Event in Real-Time?

  BacPath 13: Molecular Analysis of Bacterial Pathogens
  Tuesday 29th September, 2015, Poster Session 3

- Identification of a Unique Antibiotic Resistance and Virulence Locus from a Human Blood-Stream Infection

  International Conference on Plasmid Biology
  Poster presentation, 2016

- A uropathogenic *Escherichia coli* ST405 with a unique complex resistance region

  Mobile Genetic Element: Transposable Elements, Integrative and Plasmids
  Oral presentation number 8, 2017

Abstract

Tracking resistance genes based on specific structural features of class 1 integrons is an integral part of clinical epidemiology. A class 1 integron is a gene capture and expression unit, most frequently found in Gram-negative bacteria. They are known to be one of the greatest contributors to the spread of multi-drug resistance genes within clinical isolates.

Recent studies indicate that certain insertion elements target specific regions of class 1 integrons creating unique structures. Consequently, the resistance pool in such integrons goes unnoticed in standard molecular screening methodologies, although they are equally efficient in expressing and disseminating resistance genes. Examples of such class 1 integrons with atypical structures, now realised to have widely disseminated within human and animal *E. coli* populations worldwide, were generated by the insertion of a genetic element known as IS*26*. This project aimed to determine the presence of such structures, as well as characterise the diverse complex resistance loci (CRL) and virulence cargo of extraintestinal pathogenic *E. coli* samples in a specific Sydney hospital. This was achieved by identifying the virulence-associated and antimicrobial resistance gene pool harboured by these strains with a particular focus on insertion or deleterious molecular events, discerning the locations of CRL within genome or plasmid where possible and observing trends within clonal groups.

Targeted PCR, Sanger sequencing and next generation short and long read genome sequencing techniques in conjunction with bioinformatic analyses were used to characterise clinical *E. coli* samples from Sydney Adventist Hospital isolated between 2009 and 2011.

Our data suggests that antibiotic resistance is readily transferring between host populations via lateral gene transfer of mobile elements in Sydney through the observation of unique molecular signatures. Clonal groups were identified in the cohort which share virulence and antimicrobial resistance traits, in some cases at seemingly differing stages of evolution. Large scale studies such as this provide insight to the mechanisms and forces driving the dissemination of antimicrobial resistance.

Contents

Abbreviations

AIEC	adherent-invasive *E. coli*
APEC	avian pathogenic *E. coli*
bp	base pair
CDS	calibrated dichotomous sensitivity
CRL	complex resistance loci/locus
CS	conserved segment
DAEC	diffusely adhering *E. coli*
EAEC	enteroaggregative *E. coli*
EHEC	enterohemorrhagic *E. coli*
EIEC	enteroinvasive *E. coli*
EPEC	enteropathogenic *E. coli*
ESBL	extended spectrum β-lactamase
ETEC	enterotoxigenic *E. coli*
ExPEC	extra-intestinal pathogenic *E. coli*
GI	genomic island
GIT	gastro-intestinal tract
HC	hemorrhagic colitis
HUS	haemolytic uraemic syndrome
Inc	incompatibility
IPEC	intestinal pathogenic *E. coli*
IS	insertion sequence
kb	kilobase
LB	Luria-Bertani/lysogeny broth
LGT	lateral gene transfer
MDR	multi-drug resistant
MGE	mobile genetic element
MLST	multi-locus sequence typing

NMEC	neonatal meningitis *E. coli*
nt	nucleotide
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
SMAC	sorbitol-MacConkey agar
SMRT	single molecule real time
SNP	single nucleotide polymorphism
SRA	sequence read archive
ST	sequence type
TU	translocatable unit
UPEC	uropathogenic *E. coli*
UTI	urinary tract infection
VAG	virulence-associated gene

Chapter 1: Background

1.1 Antibiotics and the Evolution of Resistance

Antimicrobial resistance genes have been traced back to before the harnessing of naturally occurring antimicrobial compounds for medical use, with those conferring resistance to β-lactams, tetracyclines and glycopeptides having been isolated from 30000 year old permafrost[1]. It was initially believed that the original function of these compounds lay in eradication of competing microorganisms. However, it's been shown that sub-inhibitory concentrations can induce the production of biofilms, instigate SOS response and affect metabolism. Biofilms are a community of adherent bacterial cells living within an extracellular matrix and can occur on virtually any moist warm surface, including medical implants and equipment (catheters, artificial valves etc.). They are notoriously difficult for antibiotics to penetrate therefore complicating infection treatment options. The bacterial SOS response temporarily increases the basal rate of recombination and mutation events[2, 3], thus exposure to antibiotics leads to diversity amongst the bacterial pangenome. So, it is clear that the original function of antimicrobial compounds transcends their ancestry. This causes significant issues in the modern era regarding the development of new antimicrobials, antimicrobial usage and the formation or dissemination of resistance genes.

The advent of the "antibiotic era" is possibly the greatest feat of modern medicine. Allowing for the treatment of previously fatal infections, the discovery and usage of antimicrobials has not only saved countless lives on an individual case basis but when used preventatively halts the progression of potential epidemics[4, 5]. However widespread implementation of antimicrobials has placed an intense selective pressure on bacteria as a whole to overcome this risk to their survival. The plasticity of microbial genomes threatens the success of antimicrobial interventions on a global scale and is not limited to human medicine. "Multidrug resistant" is the term applied to a microorganism that has developed resistance to three or more classes of antibiotics. There are a number of causes for the evolution of MDR, the largest contributor being the inappropriate use of antibiotics[6]. For example, the overuse of drugs essentially forces bacterial communities to adapt via resistance mechanisms in the face of selective pressure, whereas under use such as incomplete antibiotic courses allow survival of the fraction of the bacterial population utilising mutations and adaptive mechanisms. After the development or acquisition of resistance

determinants, uniquely adapted bacteria can proliferate and infect subsequent hosts or, in the instance of acquired traits, propagate their newfound arsenal within the surrounding bacterial community inevitably contributing to the serious threat posed by MDR.

The threat to human medicine is perpetuated through veterinary practices as, in addition to the environment, these niches form a web of interconnectedness where all realms are significantly impacted by each other (Figure 1.1). Bacteria that develop resistance to antibiotics in an agriculture setting can be transferred to humans through handling of infected animals or contaminated meat products[7-10], a process termed zoonosis with respect to disease causing bacteria. This adds an indirect route of transmission of resistance genes between animals and the human population[6]. Historically, there have been numerous outbreaks of food-borne origin, particularly with respect to enterohemorrhagic *E. coli* serotype O157:H7 (Figure 1.2). Policies involving the restriction of medically significant antibiotics in food-producing animals, prudent and judicial use of antibiotics, the banning of growth promotion, removal of heavy metals such as zinc and copper from animal feeds, and the advocacy of non-antibiotic treatment methods such as probiotics and phage therapy[11, 12] have been put forward as a means of controlling the spread of resistance, but have only recently been stringently regulated or imposed[6].

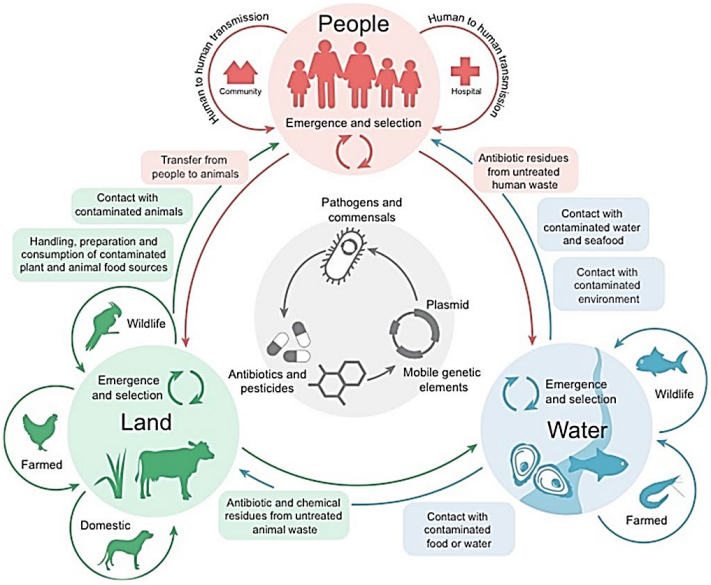


Figure 1.1: Visual representation of the transmission routes of bacteria and mobile genetic elements between human medicine, veterinary/agricultural practices and the environment, as well as the selective pressures exerted in each niche (Ausgem website).

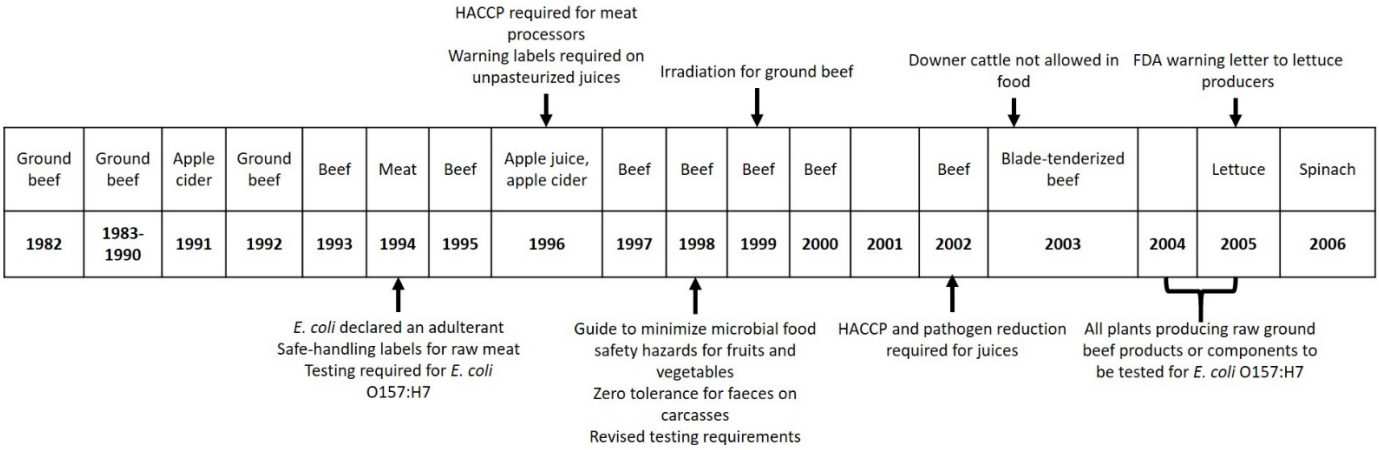


Figure 1.2: Timeline of *E. coli* O157:H7 outbreaks of food borne origin in the US between 1982 and 2006, showing new regulations in response. Adapted from Doyle *et al*. 2006[13].

Irrespective of the host and way MDR infections arise within any population, it imposes massive economic burden on agriculture, veterinary practices and the healthcare system in the form of increased costs of diagnosis and treatment[14]. For example, a 2012 study established that MDR *Pseudomonas aeruginosa* costs 70% more per admission than non-resistant strains[15]. It is for this reason that an in depth understanding of the ramifications and the reasons for such rapid dissemination of bacterial resistance is required, and widespread awareness is key.

The hospital/aged care environment and anthropogenic waste are major contributors to the evolution and transmission of antibiotic resistance due to the opportunity for bacteria from diverse genera to interact, both with each other and pollutants such as unmetabolized antibiotics and disinfectants. Multi-drug resistant bacteria have been proven to transmit via human contact (patients and hospital staff), medical equipment, bedding and fomites[16, 17] as well as form biofilms[18, 19] which are notoriously difficult to eradicate. Antibiotics excreted by hospital patients and members of the community undergoing treatment are released into wastewater and from treatment plants can permeate environmental niches via use as fertiliser or entering runoff. It is believed that the use of disinfectants in hospitals promoted the rise in incidence of class 1 integrons; immobile gene expression elements containing a site-specific recombination system which can uptake free gene cassettes that may confer resistance to a number of antibiotics. Class

1 integrons are a reliable proxy for multi-drug resistance, as they contain the quaternary ammonium compound resistance gene *qacEΔ1* in their typical structure.

Over the last century, humans have increased the use of antimicrobials significantly thereby inducing the spread of resistance through strong selective pressure and winnowing bacterial populations down to those capable of expressing a resistant phenotype. Antibiotic resistance in any bacteria can occur via two mechanisms, intrinsic and acquired, each contributing heavily to diversity amongst bacterial genomes. Intrinsic refers to naturally occurring resistance or mutations in chromosomal genes with an original function unrelated to antimicrobial resistance. Acquired resistance is the acquisition of resistance genes from the global resistome by a genetic phenomenon called lateral gene transfer (LGT). DNA sequencing has revealed that the resistome is older and larger than its current subset of clinically important antimicrobial resistance genes[20], with these genes having originated in natural environments such as aquatic sources and soil; for instance, multi-drug resistant bacteria dating back 4 million years[21] have been isolated from underground, and the extended spectrum β-lactamase gene family CTX-M originated in the chromosome of soil/water dwelling *Kluyvera* species[22].

Human selective pressures forced recruitment of resistance determinants onto mobile elements including plasmids, transposons and insertion elements leaving them open to propagation via LGT, which is arguably more detrimental to the global healthcare system than intrinsic resistance. In 2009, it was estimated that between 1.6 and 32.6% of genes in the collective bacterial genome were acquired by LGT[23]. Through the spread of genes encoding antimicrobial resistance and virulence determinants, LGT has promoted the development of pandemic bacterial clones.

1.2 Lateral Gene Transfer

1.2.1 The Concept

Lateral gene transfer is a two-step genetic phenomenon, which includes the non-genealogical transfer of DNA from a donor bacterium to a recipient bacterial cell[24], followed by integration of the acquired DNA into the genome for stable inheritance to subsequent generations. Because this

can occur between bacteria of different species, LGT is considered to be a major vehicle of bacterial evolution and genetic diversity[23]. LGT is a phenomenon by which bacteria can access a communal pool of resistance genes, referred to as the 'resistome'[25].

There are three main mechanisms by which LGT can occur- transformation, transduction and conjugation (Figure 1.3).

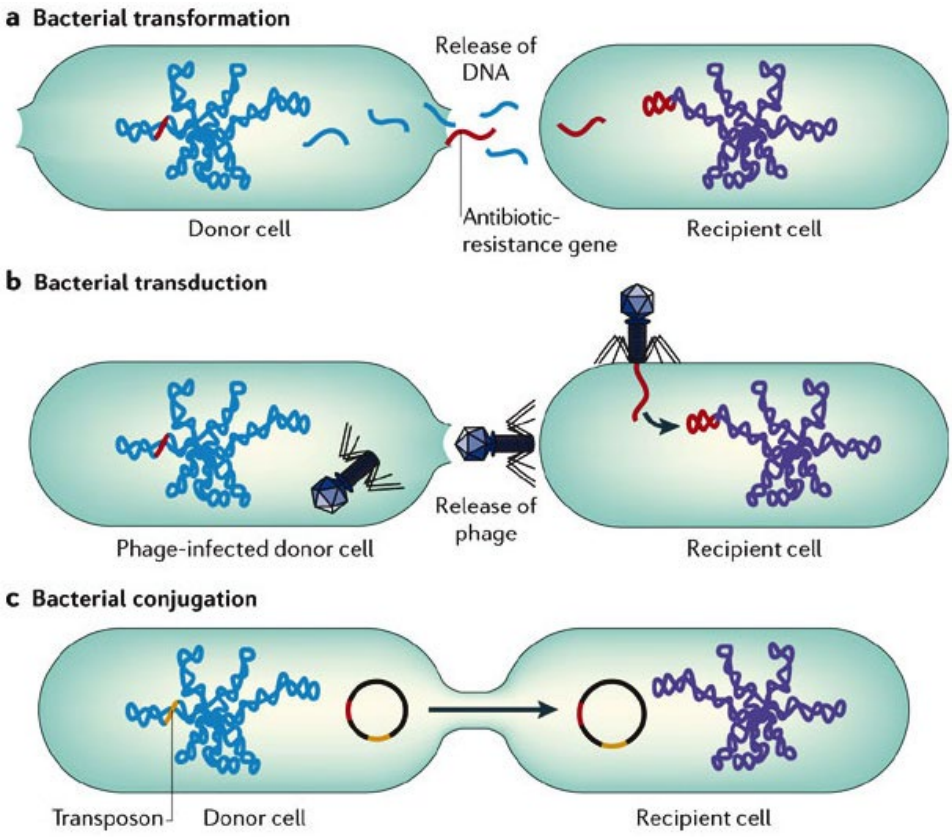


Figure 1.3: Graphic of the three main mechanisms of LGT[24].

Transformation occurs when naked DNA released into the environment by lysed or damaged donor bacteria is acquired by a recipient cell and incorporated into the genome via homologous or site-specific recombination. This method can potentially transfer DNA between organisms that are distantly related[26], as the success of uptake depends on the stability of free DNA in the environment and the ability of the recipient bacteria to be naturally competent. The newly acquired DNA must provide an evolutionary advantage to the recipient bacterium in order to be stably inherited[25].

Transduction involves phage-mediated transfer of DNA to a recipient bacterium. The replication of a phage requires the packaging of phage DNA into capsids, host bacterial cell lysis and subsequent release of the newly synthesized phage into the environment where they can locate and infect recipient bacteria[26]. This process does not require genome similarity between donor and recipient and has the furthest range due to phage mobility[27]. Even though phage proteins and enzymes promote incorporation of the DNA fragments into the chromosome and protect them from host restriction endonucleases[26], success of transduction depends on the survival of the host cell after phage infection, where survival rate is estimated to be approximately 20%[28]. Phages do not always package DNA efficiently and occasionally capture DNA that resides adjacent to that of the phage. This process allows foreign, non-phage DNA to be packaged into the phage capsid, representing an important mechanism for the lateral transfer of DNA.

Conjugation is the most commonly occurring LGT mechanism[29] and the associated units of transfer are plasmids. Unlike transformation and transduction, this requires the donor and recipient to be in close proximity but is not limited to bacteria, as inter-domain transmission is possible depending on the host range of the plasmid[26]. A pilus is formed to connect the two cells and a conjugation pili is extended. The plasmid is nicked, and a single strand breaks off and passes to the recipient cell where a complementary strand is synthesised by DNA polymerase. It then re-circularises and replicates, or in some cases is integrated into the recipient chromosome by either homologous recombination (requires genome similarity) or transposition (does not require similarity)[27].

1.2.2 Homologous Recombination

Homologous recombination involves a double strand reciprocal exchange of genetic material between two DNA molecules that exhibit 25-200 base pair regions of extensive similarity[30]. This process transcends single bacteria, often occurring between different strains and species. It is most successful when the donor and recipient cells are related[26], however it has been observed that unrelated or dissimilar elements can undergo homologous recombination if they both contain copies of the same translocatable element, such as an insertion sequence[30]. In terms of LGT, this recombination process is most commonly seen in association with transformation and conjugation[27].

1.2.3 Site-specific Recombination

Site-specific recombination facilitates the movement of mobile genetic elements between non-homologous (dissimilar) regions in a genome, either on a single chromosome or between chromosomes[31]. Examples of such recombination systems include phage-mediated integration and the recruitment of gene cassettes into integrons.

The bacteriophage integrase involved in phage-mediated integration belongs to one of two protein families; serine recombinases, which perform double stranded breaks in DNA, or tyrosine recombinases, which generate a single stranded break[32]. Phage integrases target recombination site pair *attP* and *attB* for integration (less than 50 base pairs), and *attL* and *attR* for excision[33].

The process by which gene cassettes are recruited into an integron occurs by integrase-mediated integration. Depending on the orientation of the recombination site/s involved, in the case of integron gene cassette systems, it can result in the insertion of an element or its excision for translocation to a new site. In bacteriophages it may lead to inversion of the element, which in the case of genes can lead to them being "switched off"[31]. Such a recombination process can also give rise to spontaneous mutations, which it often does[31].

1.3 Established *Escherichia coli* Lineages

Escherichia coli are facultative anerobic Gram negative bacilli residing amongst the commensal gut microbiota of all warm-blooded organisms, as well as some reptiles[34]. In 2008 it was estimated the *E. coli* pangenome exceeds 13000 genes, however it is currently referred to as having an "infinite pangenome"[35, 36]. The same 2008 study calculated that the core genome comprises 2344 ± 43 genes[37] indicating that lateral gene transfer plays an important role in its evolution. Belonging to the family Enterobacteriaceae, *Escherichia coli* are generally benign in the gastro-intestinal tract (GIT). However, these pathogens are adaptable to a broad range of environments. Resilience and survival are achieved through the acquisition of pathogenicity determinants such as antimicrobial resistance genes, virulence associated genes and metabolic

genes[38], as well as the evolution of point mutations. *E. coli* pathotypes, determined by accessory gene cargo acquired via mobile elements, can be placed under two umbrella categories based on their method of infection, intestinal and extra-intestinal. Intestinal pathogenic *E. coli* (IPEC) for example enterohaemorrhagic *E. coli* (EHEC), tend to infect via the fecal-oral route and result in symptoms such as diarrhoea and haemolytic uraemic syndrome due to direct pathogenesis of the GIT and subsequent release of Shiga toxins into the circulatory system. Toxins are important in the pathogenic arsenal of IPEC, as they divide the group further into sub-pathotypes; enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), enterotoxigenic *E. coli* (ETEC), enterohemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), diffusely-adherent *E. coli* (DAEC), and adherent-invasive *E. coli* (AIEC)[39]. Extra-intestinal pathogenic *E. coli* (ExPEC) on the other hand cause infection outside the gut, with the most common site being the urinary tract. Urinary tract infections can often occur via self-infection of host commensal bacteria due to poor hygiene, sexual intercourse and during child birth, or any situation that establishes a direct transmission route for fecal bacteria to the urethra[40]. Infections of the urinary tract are commonplace and account for widespread morbidity and mortality; in 1997 it was estimated that UTI resulted in 7 million doctor's office visits and 1 million emergency department visits in the US alone[41]. Uropathogenic *E. coli* (UPEC) can ascend the urinary tract and colonise the upper urinary tract causing infections on the kidneys (pylonephritis) and blood stream infections (sepsis). Antibiotics are essential for the treatment of these life-threatening infections.

Highly virulent and antibiotic resistant UPEC strains are emerging. Specific strains and lineages are major contributors to the problematic treatment of UPEC, with studies suggesting between 10 and 20% of all urinary tract infections could be caused by a subset of clonal groups or sequence types (ST)[42, 43], most notably ST69[42, 44, 45], ST73[46, 47], ST95[42], ST405[48, 49] and ST131[50-53]. ST131 is the most well studied of these sequence types, and is considered a major concern for public health, being commonly associated with the carriage of genes encoding ESBLs and genes encoding resistance to trimethoprim-sulfamethoxazole and fluoroquinolones[54-56]. This sequence type was first identified in 2008 and was observed in nine countries over three continents[57, 58]. It was surmised that the emergence if this lineage occurred over less than a decade[56, 59], but resulted in ST131 becoming the most dominant ExPEC clonal group globally and a "major etiological agent" of urinary tract infections, pyelonephritis and urosepsis[52]. ST131 consistently exhibit a set

of virulence factors including the uropathogenic specific protein *usp*, the fimbrial adhesion *fimH*, the secreted autotransporter toxin *sat*, the aerobactin receptor *iutA* and the pathogenicity marker *malX*[53]. The speed at which this clonal group has disseminated globally highlights the threat that can be caused by a single clonal group as well as the need to monitor emerging clones. Although it has been demonstrated that ST405 generally harbour more antibiotic resistances than ST131[49], literature focussing on this emerging sequence type is scarce with the vast majority of studies into ExPEC clonal groups focussing strictly on ST131. Recent research is opening up to other established and emerging ExPEC lineages[60-62], however studies are still few and far between by comparison.

This prejudice in reporting so-called dominant sequence types exists because molecular assays are often developed to rapidly screen for what are considered to be important pathogens or high-risk antimicrobial resistance genes. Ease of use and widespread application of diagnostic tests allows researchers and clinical laboratories to report the frequency of a particular clone detected by the targeted assay and ignores other emerging sequence types. This, as well as the lack of literature on emerging clones, represents a bias towards established strains over those which are emerging and potentially highly antibiotic resistant. This scenario is not unprecedented in modern clinical microbiology. It could be argued that the importance of non-O157 enterohaemorrhagic *E. coli* (EHEC) as a significant cause of haemolytic uremic syndrome (HUS) and haemorrhagic colitis (HC) in the USA was not fully appreciated during early investigations of the causes of these diseases[63, 64]. This was likely due to several reasons including the prevailing dogma that O157:H7 EHEC was responsible for causing HUS/HC. The diagnostic test to detect O157:H7 EHEC was simple and cost-effective because it exploited the observation that O157 EHEC do not ferment sorbitol. The development of sorbitol-MacConkey-agar (SMAC-agar) where sorbitol replaced lactose was important in routine surveillance for O157:H7 EHEC[65]. SMAC-agar was widely adopted in clinical laboratories around the globe because it was inexpensive and effective. Clinically significant non-O157 EHEC serotypes ferment sorbitol and diagnostic tests to detect them are more complicated and were not always routinely implemented[64]. More recent estimates of the causes of human EHEC infections suggest that between 30-60% of cases are caused by non-O157 EHEC[63]. In Europe EHEC O26 is a leading cause of human EHEC infection[66] but others including shiga toxigenic *E. coli* serotypes

O111:H8, O26:H11, O103:H2, O45:H2, O121: H19, O145:H28, and O157:H7 are recognised as leading causes of EHEC infections globally[63].

1.3.1 Lateral Gene Transfer and Drug Resistance in *E. coli*

Escherichia coli are Gram-negative human and animal gastrointestinal commensals of the family Enterobacteriaceae. They are well known food-borne infectious pathogens, and they play detrimental roles in a wide variety of nosocomial (hospital associated) infections. *E. coli* are increasingly found to be multi-drug resistant. Antibiotic resistance genes show a propensity to cluster within complex resistance loci located on plasmids and genomic islands. It is not uncommon to find genes encoding resistance to extended spectrum β-lactams, which causes resistance to commonly used as well as last-line antibiotics such as penicillin, aztreonam, first generation cephalosporin and occasionally third generation cephalosporins[67]. *E. coli* and other members of the Enterobacteriaceae are exposed to antimicrobial selection pressures in hospitals, human and animal effluent facilities, food animal production, and in agriculture more broadly and these pressures are likely to profoundly influence how resistance genes assemble on mobile genetic elements. Multi-drug resistance in *E. coli* severely complicates and limits treatment options, especially with resistance to clinically important and last-line antibiotics such as carbapenems and fluoroquinolones emerging.

Although *E. coli* is found as normal flora of the human gut, there are a number of disease-causing pathotypes that have evolved over time due to LGT of antimicrobial resistance genes and virulence determinants[26]. Enterotoxigenic *E. coli* (ETEC) for example encodes for heat labile and heat stable enterotoxins that, along with colonization factors, are encoded by a plasmid acquired via LGT[67]. Similarly, enteroinvasive *E. coli* (EIEC) has acquired plasmid-borne factors via LGT which enable it to penetrate and multiply within the host intestinal epithelial cells. Shiga toxins that are important in the aetiology of enterohaemorrhagic *E. coli* (EHEC) disease are transferred via bacteriophage (transduction) that were originally described in *Shigella* species[67].

From an evolutionary perspective, LGT has widely contributed to the survival, resistance and pathogenicity of *E. coli.* The O157:H7 strain of EHEC in particular has caused numerous outbreaks of haemorrhagic colitis globally[68-72]. In the United States, it has been estimated to cause 73000 infection cases yearly, resulting in approximately 60 deaths[67], costing the economy

$405 million in 2003 alone[73]. It is projected that between 10 and 16% of the *E. coli* genome was acquired via LGT[26].

The acquisition of combinations of virulence factors, antimicrobial resistance genes and mutations that influence how effectively disease-causing organisms transmit from one patient to the next are important elements of successful pathogens. The presence of specific combinations of virulence genes determines pathotype designations. Uropathogenic *E. coli* (UPEC), which tends to harbour a distinct suite of virulence genes and factors allowing colonisation and invasion, is also increasingly found to be resistant to fluoroquinolones[50], trimethoprim-sulfamethoxazole, aminoglycosides, tetracycline[74] and extended-spectrum cephalosporins. These antibiotics are commonly used in the resolution of urinary tract infections, so treatment options have become exceedingly limited, in turn detrimentally affecting associated morbidity and mortality rates[52]. The contribution of LGT to the dissemination of resistance in UPEC is exemplified by the current pandemic of CTX-M genes, found on mobile genetic elements[52] and are therefore able to spread rapidly. CTX-M genes confer resistance to extended-spectrum β-lactams, with the CTX-M-15 variant currently representing the most widespread CTX-M enzyme globally[57, 58].

1.4 Mobile Genetic Elements

Mobile genetic elements (MGE) are DNA sequences of varying lengths, which can mobilise between replicons and capture new genetic cargo, thus leading to reorganisation of genes[75-77]. They may carry the functional components required for independent transmission and integration into a host genome, in the form of DNA segments coding for necessary proteins and enzymes[78]. In addition to this, they can carry a suite of accessory genes including but not limited to those conferring resistance to antimicrobials or genes for colicin immunity, which provide distinct selective advantages to a bacterial cell and its progeny in terms of evolution and survival. MGE include plasmids, transposable elements, gene cassettes and under certain circumstances, integrons.

1.4.1 Plasmids

Plasmids are primarily extra-chromosomal double-stranded circular DNA molecules, can be capable of transformation or conjugation and are recognised as the major vehicles of LGT[29]. Since plasmids self-produce the proteins necessary for replication initiation, they can replicate via conjugation (Figure 1.4) through association with host proteins and enzymes[79]. Replication begins at the plasmid origin, or *ori* site, and is initiated by the binding of the replication protein encoded by the *rep* gene. Together the *ori* site and *rep* gene are the minimum required components for the plasmid to replicate, with the remainder of the replication machinery provided by the host chromosome[80]. In Gram-negative bacteria, a type 4 secretion system (T4SS) physically bridges the donor and recipient cells via a conjugation pilus, allowing for transfer of single-stranded DNA which is then synthesised into a double stranded plasmid within the recipient. Mechanisms of plasmid transmission between cells affect the host range of the plasmid, which may be "narrow" or "broad", dependant on the number of bacterial species in which the plasmid survival can be maintained.

Plasmids are frequently observed carrying other mobile genetic elements, such as complex antimicrobial resistance loci formed by transposons and insertion sequences, greatly contributing to their dissemination and consequently the spread of resistance. Plasmids inflict a metabolic burden on host cells resulting in decreased growth rate[81] that is generally tolerated due to their significant contribution to the survival of the bacterium under antimicrobial stress and selection pressures, or in specific environments and bodily sites where their genetic cargo provides a selective and competitive advantage to the host bacteria that carries them. Nonetheless, large conjugative plasmids are tolerated and have persisted; compensatory mutations are thought to play a significant role in alleviating that fitness impost. Fitness cost and metabolic burden can be minimised through copy number control, for which there are two methods; transcription repression via antisense RNA and binding together of *ori* sites which minimises replication. Due to larger plasmids requiring a lower copy number in order to be maintained, they generally harbour stability, partitioning and postsegregational killing systems[82] to reduce chances of being outcompeted. Partitioning systems exist to distribute plasmids between daughter cells, whereas postsegregational killing systems ensure death of those daughter cells which do not inherit the plasmid and act as a failsafe to partitioning system malfunction.

Despite any metabolic burden, under certain circumstances plasmids provide an evolutionary advantage that assists in ensuring the success of the cell. Plasmid fitness studies have mostly been conducted under *in vitro* conditions where bacteria are undergoing exponential growth. Because of the difficulties in replicating how wildtype bacteria live there is a perception that the fitness impost afforded by plasmids may be over emphasized. There is even evidence that some plasmids provide a fitness advantage to their host, however the intricate details of plasmid fitness cost are not well understood and may vary niche to niche. *In vivo* bacteria form biofilms, and within these structures bacteria belonging to the same lineage (i.e. different clonal isolates) may have different metabolic burdens depending on where they reside within the biofilm, with different capacities to carry plasmid cargo. More studies are needed to understand plasmid fitness[83].

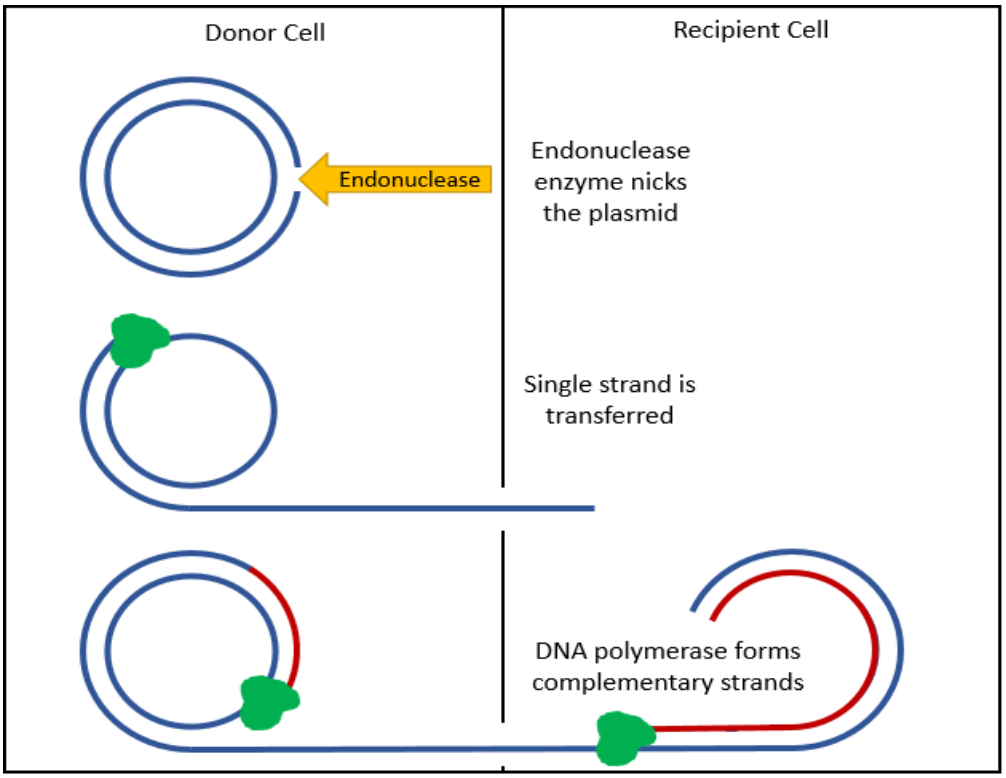


Figure 1.4: Conjugation of a plasmid.

Plasmids can be categorised based on their replication systems. This is termed "incompatibility type" and refers to the inability for two closely related plasmids to exist stably within a cell if their replication machinery interferes with one another, affecting copy number control and consequently stability. There are several main incompatibility types circulating within

Enterobacteriaceae, with F-type plasmids, encompassing several Inc types, being the most prominent. It is important to not however, that F-type replicons are generally capable of coexisting within the same cell. Further to this, incompatibility types can be subcategorised based on the sequence of their replication genes resulting in a "plasmid sequence type" which can be used to monitor spread of a specific plasmid backbone[84]. This is referred to as pMLST and, while still baring some limitations based on the number of gene fragments used in assignment of an ST, is more effective in understanding plasmid relatedness and diversity than Inc type alone.

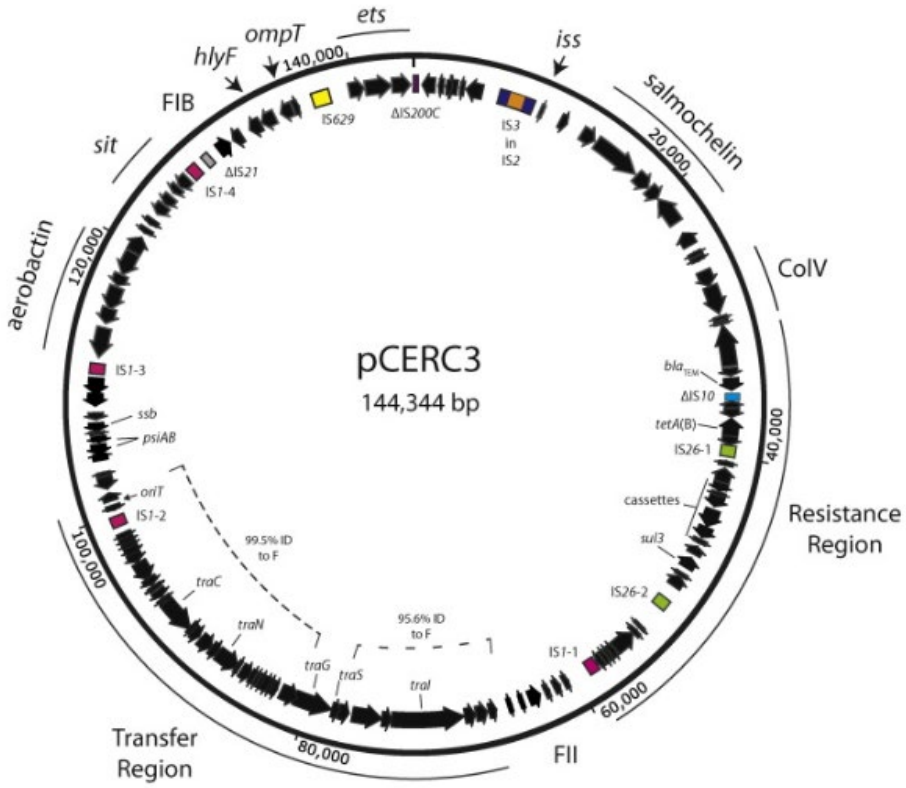


Figure 1.5: Example of a plasmid containing numerous virulence associated and antibiotic resistance genes from a commensal *E. coli* ST95 strain[85].

1.4.2 Transposons and Insertion Elements

In clinically important bacteria, transposable elements generally mediate the movement of antimicrobial resistance encoding genes between similar and dissimilar replicons. The cargo of accessory genes that transposons carry in a clinical context are almost invariably antimicrobial resistance genes, in addition to genes encoding necessary transposition functions. They are independently mobile units which cannot self-replicate, therefore relying on non-homologous integration into an autonomously replicating unit for stable inheritance[78]. The boundaries of this mobile unit, demarcated by inverted repeats (generally 9 to 40 bp), are recognised by the respective transposase protein as the 'mobile unit' during transposition.

Based on the modes of transposition, transposons can be divided into two classes. Class 1 transposons transpose by the "copy and paste" mechanism, whereby they create a copy of themselves via an RNA intermediate, which is then inserted into a recipient section of DNA. In contrast, class 2 transposons are excised from their location and "jump" to a new site by the "cut and paste" method[86].

Insertion elements (or insertion sequences; IS) are much more compact sequences also flanked by short inverted repeats, with the segment between these repeats coding for the transposase (*tnp*), the enzyme responsible for transposition[87]. These do not carry any accessory genes within their immediate boundaries (defined by inverted repeats) but can capture different genes between two copies of the same IS or two very similar IS as a composite transposon[88](Figure 1.6). Composite transposons effectively mobilise a single gene or group of genes and assist in gene expression by creating a promoter[87].

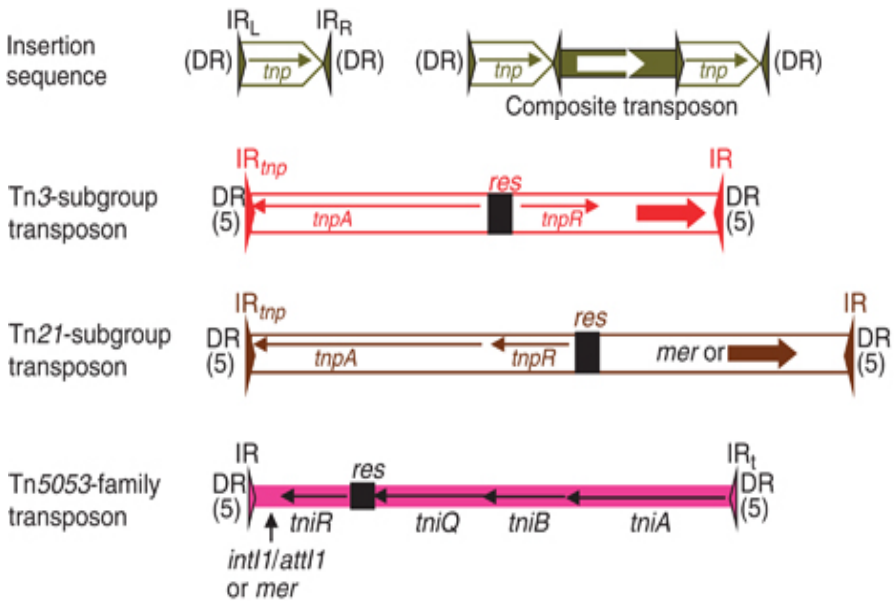


Figure 1.6: Structure of insertion elements and formation of composite transposons, as well as examples of transposon families (adapted from Partridge et al 2011[87]).

1.4.3 Integrons

Integrons are two component, generally immobile gene expression elements that contain a site-specific recombination system. There are a number of integron classes, divided and characterized by their unique integrase protein sequence of which classes 1, 2 and 3 are clinically prevalent [89]. Integrons have been reported in bacteria from varied environmental niches including soil, plants, marine environments and marine sediments[90-93] and their genetic cargo varies enormously in these environments. It was estimated that greater than 15% of all whole genome sequenced bacteria contained at least one integron[94, 95]. Location of the integron determines the degree of evolutionary advantage; the metabolic burden on the cell is lessened by a chromosomal location and can be inherited vertically, whereas plasmid housing places stress on the containing cell but the integron can be mobile and therefore be transferred to differing species. Regardless of location however, integrons provide an advantage to the host.

A class 1 integron is a gene capture and expression unit, and the most commonly encountered class of integron within clinically relevant pathogens touting a particularly high incidence within

Gram-negative bacteria[96], however observations have occurred in Gram-positive bacteria[97]. Class 1 integrons and were initially captured within a functional Tn*402*-family transposition module. Due to evolutionary changes this module is generally incapable of self-directed transposition in clinical class 1 integrons but may have its transposition machinery supplemented *in trans*[98]. Structurally, class 1 integrons are comprised of two conserved segments (CS), 5′-CS and 3′-CS, between which is located the 'variable region' made up of a gene cassette or multiple cassettes known as the cassette array. The 5′-CS includes the core integron; the integrase gene, an integrase promoter (P_{int}), an attachment site (*attI*) and a cassette promoter (P_c), while the 3′-CS, in the majority of class 1 integrons, contains a truncated quaternary ammonium compound resistance gene (*qacEΔ1*) and sulphonamide resistance gene (*sul1*), although in some cases an *orf5* gene is also present. The 3′CS is often the target of insertion sequences, resulting in unique molecular signatures that can provide epidemiological information[99]. Recent evidence suggests that class 1 integrons structure is changing predominantly as a result of the action of IS*26*[99, 100].

The main function of a clinical class 1 integron is capturing, rearranging and expressing promoterless elements known as gene cassettes[101]; thus providing a quick and flexible means of adaptation to antimicrobial selective pressures where cassettes proximal to the promoter receive a higher level of expression. In a clinical context, gene cassettes typically encode resistance to a variety of antibiotics, but in an environmental context can encode efflux pumps and proteins of unknown function. The process by which they are recruited into an integron occurs by site-specific recombination (recombination between non-homologous regions)[102]. The integrase protein catalyses recombination between the core integron and gene cassettes, which generally include a single coding sequence and a recombination site, *attC* (between 57 and 141 base pairs long)[89], although fused hybrid cassettes have been observed[50, 74, 103]. The *attI* site of the integron undergoes recombination with the *attC* site of a gene cassette, resulting in a single strand exchange and consequently integration[104]. Gene cassettes exist as covalently closed circular molecules in bacteria which linearize for insertion, forming a 'cassette array' when assembled in tandem[101]. They are promoterless and therefore cannot replicate or express independently of an integron. Depending on the orientation of the recombination site/s involved, in the case of integron gene cassette systems, it can result in the insertion of an element or its excision for translocation to a new site. Since the integrase protein specifically recognises the *attI* site rather

than the *attC* site, tandem recruitment of gene cassettes is possible at *attI* and the further downstream a cassette is from the promoter of the integron, the lower its level of expression[105].

Physical boundaries of clinical class 1 integrons are delimited by inverted repeats IR_i (at the 5' end) and IR_t (at the 3' end), characteristic of Tn*402*-family of transposons. The transposition module also contains the putative transposition genes *tni*R, -Q, -B and –A as well as a resolution site[106]. However, the transposition module found in the majority of clinical samples is truncated, therefore class 1 integrons are defective transposons and hence immobile, although some are still reported to be functional[11, 106, 107]. This is important because class 1 integrons have become integrated into mobile elements such as Tn*3* family mercury resistance transposons, a process that ensures these elements are widely distributed in clinical and veterinary settings, where antibiotic pressures are greatest. A feature of transposons belonging to the Tn*402*-family is that they target resolution (res) sites within plasmids and transposons, which facilitate movement once the integron is incorporated[106]. In the absence of functional Tn*402* transposition modules, mobility of class 1 integrons relies solely on capture by plasmids, transposons and insertion elements or supplementation of transposition functions *in trans* by other transposons belonging to the same family[106, 108, 109]. An example of this is Tn*21*, a mercury resistance transposon which has captured Tn*402* outside of the *res* site therefore retaining its mobility. Tn*402* has a truncated *tniB* gene and therefore cannot transpose without supplementation by another transposon of the same family (Figure 1.7).

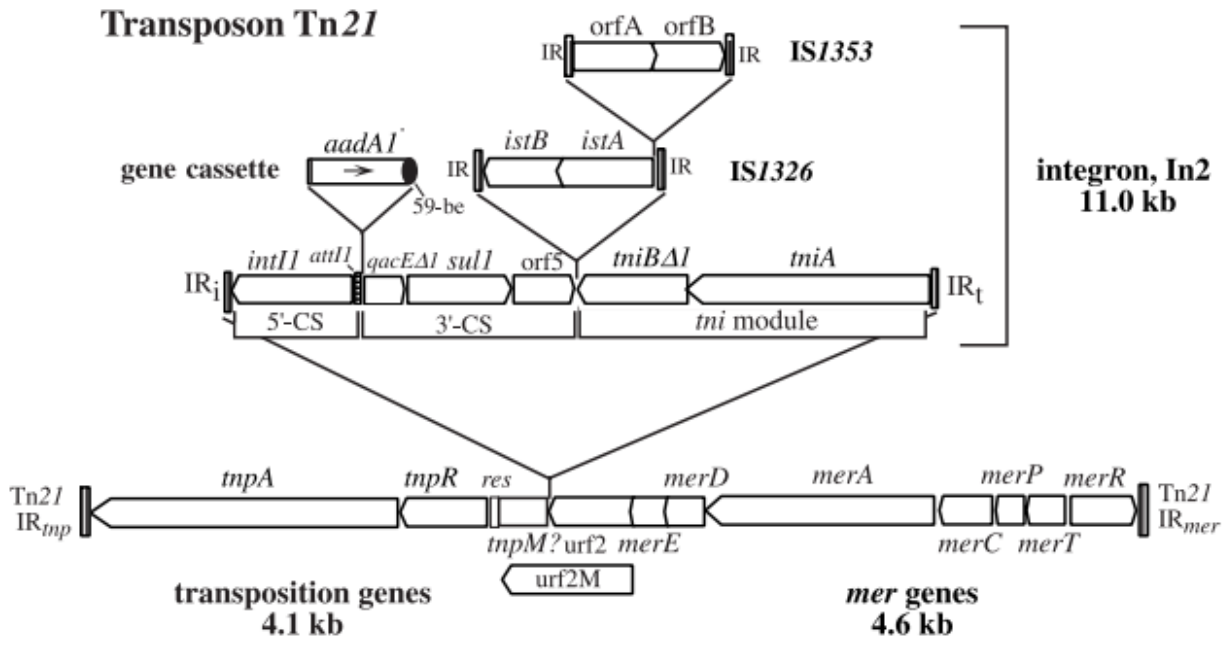


Figure 1.7: Tn*21* having captured an In2 integron which includes non-functional transposon Tn*402*[109].

1.4.4 Genomic Islands

Genomic islands (GI) may or may not be independently mobile and are often inserted in a genome at tRNA genes by site-specific recombination; differentiated from the rest of the genome by their guanine/cytosine content, codon usage and tetranucleotide frequencies as well as their tendency to contain more 'novel' genes[110]. They are highly variable regions of the genome, typically between 10 and 200 kilobases, and contain a large number of genes and other mobile genetic elements[26]. Excision from the bacterial genome requires recognition of the short direct repeats (16-20 base pairs) which flank the island at both ends[26]. GI's are named pathogenicity, symbiosis, metabolic, fitness or resistance islands based on the nature of their contained genes[110]. They are recognised as providing a significant selective advantage to the host due to their ability to encode resistance, virulence, catabolism, enhanced transmission/colonisation and many other functions which can turn a benign pathogen into a "superbug"[111].

1.4.5 Complex Resistance Loci

Mobile genetic elements have a propensity to "stack" forming complex antimicrobial resistance loci (CRL), that is a structure comprised of numerous individual units that can mobilise individually or as one locus. The locus itself can incorporate a combination of transposons,

integrons, insertion sequences and composite transposons (Figure 1.8), and while it is possible to be housed chromosomally in islands such as Salmonella Genomic Island 1[112], these CRL are more frequently observed in a plasmid context. Insertion of CRL forms direct repeats upon insertion, however these can be altered or removed by further recombination events. Insertion sites and repeats allow for epidemiological tracking of specific CRL[61, 99, 113].

Clustering of resistance is an enormous problem in the face of infection control and treatment, and only compounds the need to characterise the elements that mobilise CRL and track their spread.

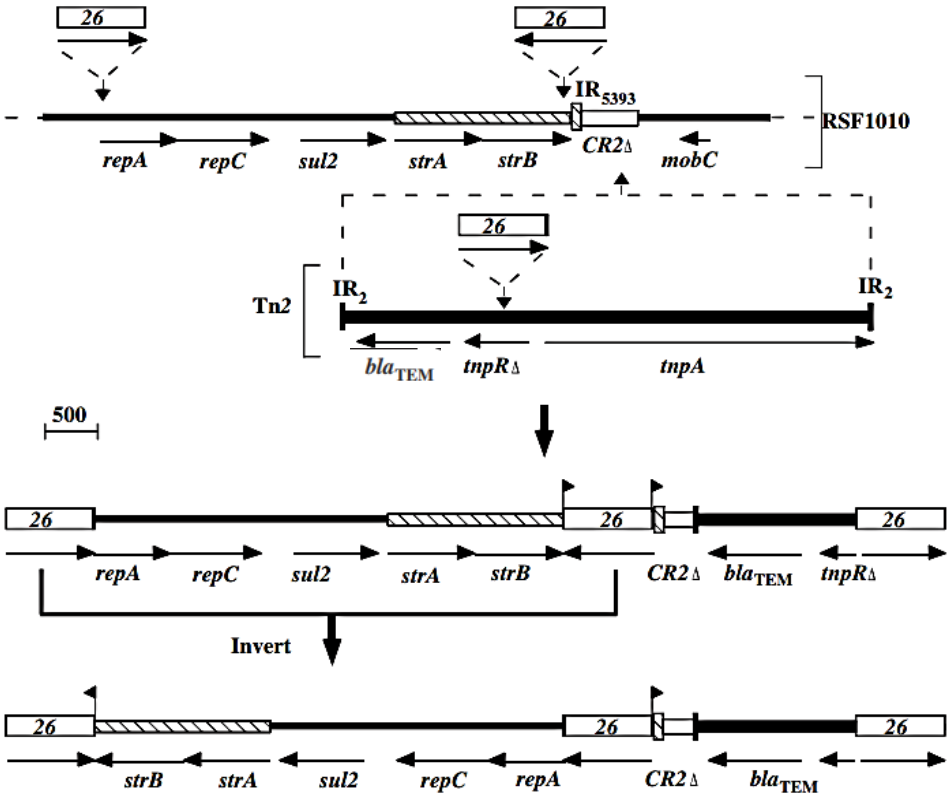


Figure 1.8: Example of a complex antimicrobial resistance loci formed by the stacking of multiple mobile genetic elements mediated by IS*26*[75].

The insertion element IS*26* appears to be one of the main drivers of antimicrobial resistance gene clustering and rearrangement, with recent studies into *E. coli* collections showing a 98% incidence of IS*26* within class 1 integron positive pig and human sourced strains[100, 114] and examples of IS*26* mediated formation of composite transposons[61, 75, 99, 115]. IS*26* is an 820 bp

insertion element composed of a single reading frame that encodes the transposase protein and is bounded by 14bp inverted repeats[87]. It is well characterised as having the propensity to form composite transposons, like Tn*6026*, Tn*6029* and Tn*4352*[116]. Most transposable elements create direct repeats of varying size upon insertion into a genetic location, and for IS*26* these are 8bp and are left upon excision. However, IS*26* has the ability to cause deletions and inversions of adjacent sequence (including direct repeats)[87] and therefore plays a critical role in shaping CRL. Tn*6029* for instance is the result of multiple IS*26* insertion, deletion and inversion events forming a final structure conferring resistance to three antibiotic classes, with segments of genetic sequence captured from plasmid RSF1010 and transposons Tn*5393* and Tn*2*.

IS*26* works via a replicative mechanism which can result in deletion or inversion of adjacent sequence[117]. A single copy can mobilise DNA via formation of a translocatable unit (TU) (Figure 1.9) and transpose via a conservative or replicative mechanism[118]. The conservative mechanism requires a secondary copy of IS*26*, whereas replicative results in a second copy.

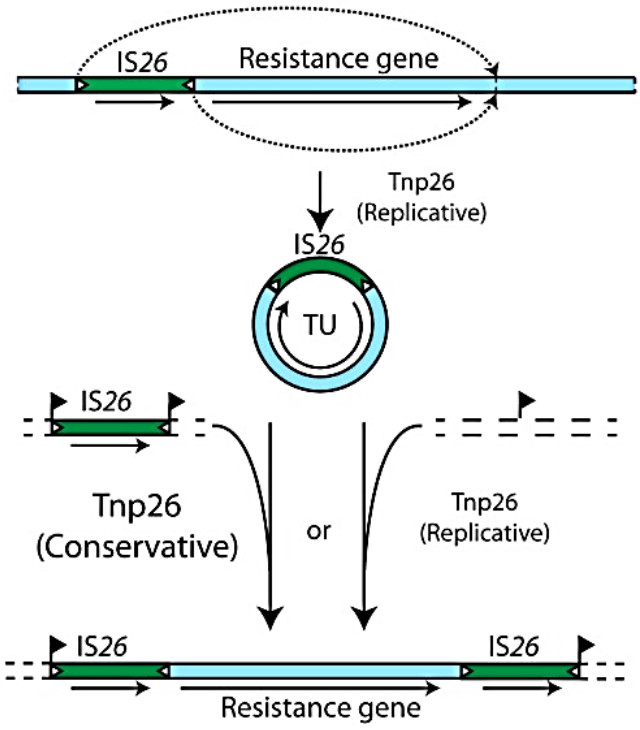


Figure 1.9: The transposition mechanisms of IS*26* adapted from Harmer and Hall 2016[118].

The clustering of antibiotic resistance determinants is not limited to IS*26* however. For example ISCR1 has been known to play a role in the capture of a range of resistance genes from their natural environments[119, 120]. The macrolide resistance operon *mphR-mrx-mphA* has also been observed in close proximity to class 1 integrons[62, 121], and this operon is associated with insertion element IS*6100*, a close relative of IS*26* with both belonging to the IS*6* family.

1.5 Class 1 Integrons: Evolution and Zoonosis

The class 1 integron originated from freshwater and soil[122], and a variant of this ancestral integron eventually became widespread in pathogens and commensals. Class 1 integrons frequently carry cassettes conferring resistance to clinically relevant antibiotics, a feature not observed in ancestral integrons but which has come about since the global implementation of antibiotic use.

One hypothesis suggests that the ancestral form was located chromosomally in non-pathogenic *Betaproteobacteria* before being captured by a functional Tn*402* transposon (Figure 1.10), carrying primarily cassettes of unknown function[123]. The *sul1* gene conferring resistance to sulphonamide antibiotics inserted into the quaternary ammonium compound resistance gene cassette *qacE* causing a deletion and forming what is now the class 1 integron 3´CS. Further deletion events occurred rendering the Tn*402* transposon non-functional and resulting in immobility while the integron itself began capturing diverse gene cassettes, resulting in the contemporary form of the clinical class 1 integron. This form can be captured by other transposons and plasmids, contributing to its dissemination and therefore the spread of multi-drug resistance phenotypes.

One such transposon, Tn*21*, is frequently found associated with class 1 integrons. Tn*21* contains an operon encoding resistance to mercury, a globally distributed heavy metal holding its origins in geothermal environments billions of years old, solidifying the ancient origins of antimicrobial resistance. Mercury resistance is advantageous to bacteria as it is widespread as a result of volcanic activity and more recently anthropogenic pollution. It and other heavy metals have been used as prophylactic measures in animal feed as well as for growth promotion[12]. In addition to this mercury resistance is often associated with antibiotic resistance genes including those conferring β-lactamase and tetracycline resistance and is therefore co-selected by antibiotic selective pressures. Tn*21* is a functional transposon in its own right but has also been known to form hybrid transposons with Tn*1721* and undergo insertion and deletion events at the hands of insertion elements, especially when associated with class 1 integrons.

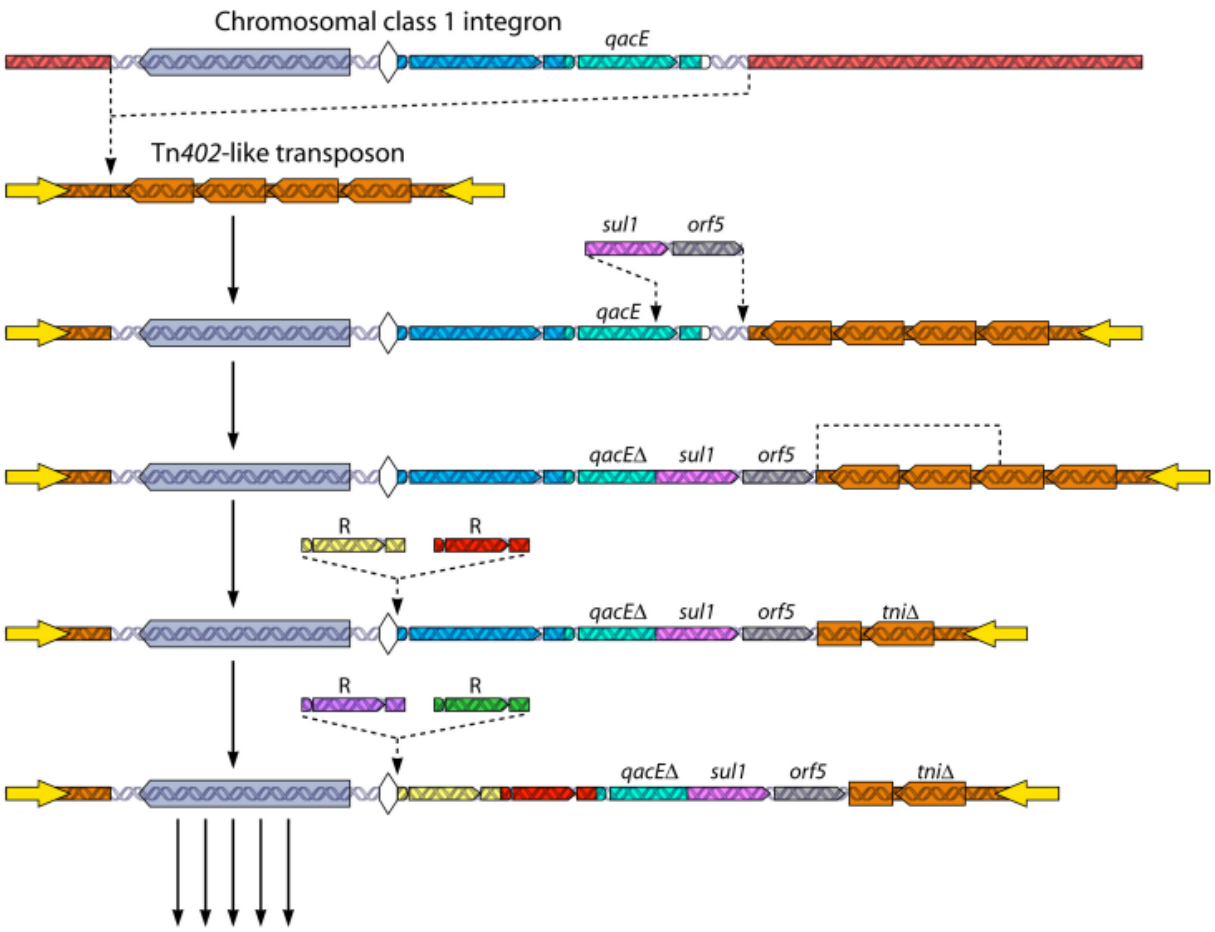


Figure 1.10: Proposed model of evolution of contemporary clinical class 1 integrons from Gillings 2014[122], beginning with the ancestral form and ending with the contemporary clinical class 1 integron.

The carriage of class 1 integrons is widespread in the Enterobacteriaceae and in other important nosocomial pathogens, such as *Pseudomonas aeruginosa* and *Acinetobacter baumannii*, and their presence is a reliable proxy for multiple drug resistance[124]. The emergence of the class 1 integron into clinically important Gram-negative bacteria is likely a response to the selection pressures afforded by the widespread use of disinfectants and antimicrobials[123].

The current form of the clinical class 1 integron is often a site for further insertion and deletion events. A significant number of class 1 integrons are atypical in that they have an altered 3′-CS[99, 106, 107].

Tracking of atypical class 1 integron-containing resistance regions has been demonstrated through the use of unique signature insertions, frequently but not solely mediated by IS*26*[99]. Once IS*26* inserted in close proximity to a class 1 integron an opportunity arose for other resistance genes, flanked by IS*26*, to cluster next to the integron and these events have undoubtedly played an important role in the clustering of resistance genes on mobile genetic elements. While IS*26* does not show any site specificity it has been demonstrated to preferentially insert near other copies of IS*26*[76]. Unique deletion events associated with the behaviour of IS*26* provides opportunities to track CRL. One example of this is the formation of an atypical class 1 integron that carries the trimethoprim resistance gene *dfrA5* and contains only 24 bp of the 3′CS due to insertion of IS*26*. This atypical integron is often associated with IS*26*-flanked transposons Tn*6029* and Tn*6026* and widely disseminated on diverse plasmid backbones in Australia in intestinal pathogen *E. coli* in bovines[99, 115] and humans[125], in human ExPEC[61], avian pathogenic *E. coli* (APEC) in Australian poultry[126] and in commensal *E. coli* from Australian swine[100].

The observation of molecular signatures in addition to crossover of sequence types puts a spotlight on cross species transmission of highly pathogenic strains and/or plasmids, and the antimicrobial selective pressures that drive their propagation.

Due to decreased cost and established methodologies microbial genomics has advanced to the point where it is now commonplace to undertake whole genome sequencing in lieu of PCR based strategies, and for this reason the number of studies reporting atypical class 1 integrons are steadily increasing. It is becoming increasingly easier to determine the context of class 1 integrons, although challenges remain. As the cost for long read sequencing becomes more affordable, our understanding of the evolution of class 1 integrons and complex resistance loci will escalate resulting in valuable epidemiological data imperative in the fight against antimicrobial resistance.

1.6 Experimental Rationale and Approach

Tracking resistance genes based on specific structural features of class 1 integrons is a fundamental consideration of molecular clinical epidemiology. They are the biggest contributors to the spread of multi-drug resistance genes within clinical isolates and are reliable markers for multiple antibiotic resistance.

Recent studies indicate that insertion elements such as IS*26* targeting specific regions in close proximity to class 1 integrons create unique structures. The propensity for IS*26* to cause deletions adjacent to the insertion site creates novel sequence signatures that can be used to track associated CRL. Many examples of class 1 integrons with atypical structures have been shown to be widely disseminated within human and animal sourced *E. coli* populations worldwide, and many of these were formed by deletion events that arise as a result of the insertion sequence IS*26*. IS*26* also produces unique deletion signatures in other important genes including *intI1* and *mefB* which adds to their epidemiological significance as markers for tracking CRL containing class 1 integrons[100].

The potential for IS*26* in tracking complex resistance loci is well documented, but little has been done in respect to testing this potential. Since Dawes published the first diagnostic assay to screen for atypical class 1 integrons formed by deletion events caused by IS*26*, a number of papers[11, 62, 99, 100, 115, 127] have reported atypical class 1 integrons that are embedded within CRL circulating within *E. coli* populations as well as other Enterobacteriaceae.

During my PhD Candidateship, I aimed to analyse a collection of class 1 integron positive uropathogenic *Escherichia coli* sourced from the microbiological unit at Sydney Adventist Hospital over a 3-year period between 2009 and 2011 via high throughput whole genome sequencing and bioinformatic analyses. UPEC isolates are the focus of this project as *E. coli* account for approximately 80% of all UTI and cost the Australian healthcare system $24.7 million annually, which highlights the prevalence and importance of UPEC in human disease. The genome sequence data was used to examine phylogeny, the repertoire of putative virulence factors and antimicrobial resistance genes found in these isolates and assist with the assembly of complex antimicrobial resistance loci. Structural anomalies in the 3′CS of class 1 integrons

caused by IS*26* deletions were characterised to demonstrate how CRL can be tracked in clinically important *E. coli* populations.

The clinical *E. coli* samples in this study were sourced from a variety of sites, with the majority isolated from mid-stream and catheter urine samples. The ultimate goal was to identify CRL and any associated IS*26* mediated evolutionary events, ascertain their genetic context and correlate this data with epidemiological and phylogenetic analyses. A strength of this study is that whole genome sequences of MDR *E. coli* from commensal swine populations collected at around the same time point are available in our research group. In addition, whole genome sequences of APEC from poultry with colibacillosis were also available. These sequences provided an unparalleled opportunity to investigate movement of plasmids, genetic signatures and other mobile genetic elements within NSW and across Australia. Opportunities to undertake comparative phylogenetic analyses and compare SNP profiles of MDR *E. coli* were afforded during the tenure of this study (manuscripts planned).

Information produced from this Candidateship could be useful in understanding and possibly tracing the dissemination of CRL throughout the human clinical as well as veterinary environment, highlighting the potential role that food production animals play in the evolution and spread of CRL in clinical pathogens.

1.6.1 List of Methodologies and Software

Methods used for publication purposes are detailed in their respective manuscripts (Chapters 2-6). Below is a list of methodologies applicable to the overarching project and programs/software used for bioinformatic analyses.

DNA Isolation and Whole Genome Sequencing – Genomic DNA for high throughput short read sequencing was isolated from culture using the Bioline Isolate II Genomic DNA extraction kit. Long read sequencing requires better quality DNA than commercial kits that promise fast extraction, so strains selected had DNA extracted via the phenol-chloroform method with minimal mechanical shearing and centrifugation. Short read sequencing was performed using

Illumina MiSeq and HiSeq platforms and Nextera chemistry, while long read sequencing was performed through the PacBio RS II platform using single molecule real time (SMRT) cells.

FastQC – FastQC provides quality control checks for raw sequence data produced from high throughput sequencing pipelines.

Aaron and Andrews Awesome Assembly Pipeline (A5) – An automated pipeline for *de novo* assembly of paired end short read sequencing data resulting in draft genomes[128].

Hierarchical Genome Assembly Process (HGAP) – Shotgun assembly of long single pass reads generated by PacBio sequencing via a three-step process of preassembly read mapping and trimming, assembly and consensus polishing. Good quality reads result in fully closed genomes.

Genomes from public domain – Genomes used in this study that were acquired from public domain were obtained through SRA or direct download from Genbank via BLASTn analysis of genomes sequenced personally.

Circlator – Circularisation of genome assemblies[129]. PacBio generates duplications at scaffold ends which can have minor sequence differences which can be solved by the use of this software.

Basic Local Alignment Search Tool (BLASTn) – Tool for alignment of nucleotide and protein sequences resulting in diagrammatical and nucleotide by nucleotide sequence alignment providing identity percentages and coverage[130]. Used for gene presence/absence screens as well as interrogation of the online GenBank database.

IS Finder – Online tool utilising BLASTn to align query sequences with an extensive curated database of insertion sequences and transposons, available at https://www-is.biotoul.fr/index.php.

Mauve – Software for generating whole genome comparisons (progressiveMauve) or rearrangement of scaffolded draft genomes to an appropriate reference (movecontigs)[131]. Provides a graphical representation of nucleotide similarity.

Electronic Multi-locus Sequence Typing (eMLST) – Utilises nucleotide differences in conserved genes to classify into sequence types and clonal complexes. *E. coli* is classified using the Achtman scheme, comprising 7 housekeeping genes, to assign a sequence type based on

alleles of these genes. Traditionally this method was performed via a combination of PCR and Sanger sequencing, however it can now be done *in silico*. BIGSdb[132] was used in this project.

Serotype Finder and Plasmid Finder – Online software for *in silico* serotyping based on sequence of cell surface antigens (O and H) and plasmid incompatibility type designation. Available at https://cge.cbs.dtu.dk/services/.

PhyloSift – Infers phylogenetic relationships between input genomes using alignment against a database of RNA sequences and 37 conserved prokaryotic proteins[133].

Harvest Tool Suite – A suite of core-genome alignment and interactive visualisation tools used for analysis of single nucleotide polymorphisms and phylogenetic inference, available at https://harvest.readthedocs.io/en/latest/.

FigTree – graphical viewing and editing tool for phylogenetic trees and production of publication quality figures.

BLAST Ring Image Generator (BRIG) – Generates circular BLASTn alignments using CGView and graphs GC content and skew as well as nucleotide identity[134].

SnapGene – Program for simple annotation of nucleotide sequence of 2 megabases or less. Can be used for generation of publication quality schematic figures as well as *in silico* PCR, identification of enzyme restriction sites and sequence alignment.

Tablet – High performance graphical viewer for genome assemblies for checking sequence consensus errors and read coverage[135].

Chapter 2: A draft genome of *Escherichia coli* sequence type 127 strain 2009-46

A draft genome of *Escherichia coli* sequence type 127 strain 2009-46

Aaron E Darling, [#1] Jessica McKinnon,[#1] Paul Worden,[1]Jerran Santos,[1] Ian G Charles,[1] Piklu Roy Chowdhury,[1,2] and Steven P Djordjevic[1]

[1]The ithree institute, University of Technology Sydney, P.O. Box 123 Broadway, NSW 2007, Australia; [2]NSW Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Private Bag 4008, Narellan, NSW, 2567, Australia

[#] These authors contributed equally

Abstract

Background

Escherichia coli are a frequent cause of urinary tract infections (UTI) and are thought to have a foodborne origin. *E. coli* with sequence type 127 (ST127) are emerging pathogens increasingly implicated as a cause of urinary tract infections (UTI) globally. A ST127 isolate (2009-46) resistant to ampicillin and trimethoprim was recovered from the urine of a 56 year old patient with a UTI from a hospital in Sydney, Australia and was characterised here.

Results

We sequenced the genome of *Escherichia coli* 2009-46 using the Illumina Nextera XT and MiSeq technologies. Assembly of the sequence data reconstructed a 5.14 Mbp genome in 89 scaffolds with an N50 of 161 kbp. The genome has extensive similarity to other sequenced uropathogenic *E. coli* genomes, but also has several genes that are potentially related to virulence and pathogenicity that are not present in the reference *E. coli* strain.

Conclusion

E. coli 2009-46 is a multiple antibiotic resistant, phylogroup B2 isolate recovered from a patient with a UTI. This is the first description of a drug resistant *E. coli* ST127 in Australia.

Keywords: Escherichia coli 2009-46, Genome, Sequencing

Supplementary material available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4155142/

Background

Escherichia coli infections of the urinary tract are among the most frequent infections reported in the developed world with an estimated 130-175 million cases per annum worldwide[1]. *E. coli* that cause urinary tract infections (UTI) are classified as uropathogenic *Escherichia coli* (UPEC), a subgroup of extraintestinal pathogenic *E. coli* (ExPEC). ExPEC also cause a range of afflictions including meningitis, septicaemia, and pneumonia and are genotypically and phenotypically distinct from diarrhoegenic *E. coli* (DEC)[2]. ExPEC are thought to be acquired orally via the consumption of contaminated food and are considered to be zoonotic pathogens[3-5]. The

emergence of multiple antibiotic resistance among ExPEC poses a serious health threat; antibiotics are an important treatment strategy for controlling UTI.

Multilocus sequence typing (MLST) is currently the gold standard for characterising *E. coli* causing UTI. No clear diagnostic markers are available for identifying *E. coli* causing UTI, but several sequence types (ST) including ST131, ST405, ST95, ST65, ST127, and ST10 are recognised UTI pathogens[6]. ExPEC ST127 are described as community-acquired and highly virulent zoonotic pathogens[3,6] but to our knowledge there are no genome sequences representing antibiotic resistant isolates of this emerging pathogen. Studies of *E. coli* causing UTI in Australia have focussed on characterising ST131[7,8] and serogroup O75 isolates belonging to clonal complex 14[9].

Here we describe the genome sequence of *E. coli* ST127 isolate 2009-46, a mid-stream urinary tract isolate from a 56 year old patient from the Sydney Adventist Hospital (SAN clinic) resistant to ampicillin and trimethoprim.

Methods

The isolate was supplied on a Sensi-agar plate from the SAN laboratories in Sydney, Australia. To confirm pure culture, a loopful of the isolate was streaked onto a Luria Bertani (LB) Agar plate and incubated at 37°C for long term storage in minus 80°C as a glycerol stock. A single colony was picked from the plate and subcultured in 10 mL LB broth at 37°C overnight. To prepare the glycerol stock culture 7 mL of the overnight was used, and genomic DNA was prepared from the remaining 3 mL. Genomic DNA for sequencing was prepared using the ISOLATE II gDNA extraction kit from Bioline.

Genome sequencing

DNA was quantified using Qubit flourometry and 0.5 ng of gDNA was used as template to construct the sequencing library, using the Illumina Nextera XT library preparation protocol following the manufacturer's instructions. However, the "PCR Clean-Up" and "Library Normalization" steps were omitted and size selection was instead performed by running balanced and pooled samples in a 1% agarose gel and excising the 600 bp to 1200 bp region of interest. The DNA was then purified from the agarose using Promega's Wizard SV Gel and PCR Clean-

Up System. Finally, an Agilent 2100 Bioanalyzer, with a High Sensitivity DNA Kit, was used to quantitate the pooled DNA library before loading onto the MiSeq with other multiplexed samples. Two MiSeq runs were carried out, one with paired-end 250 nt reads on MiSeq V2 chemistry and another with paired-end 300 nt reads on V3 chemistry. The first library was found to have an average insert size of 368 +/- 157 nt, while the second library had inserts with an average size 497 +/- 118 nt.

Assembly and annotation

The genome was assembled using the A5-miseq pipeline, a version of the A5 pipeline[10] that has been revised to process reads up to 500 nt long. Briefly, the A5-miseq pipeline consists of five stages: (1) read quality filtering and error correction, (2) contig assembly, (3) permissive draft scaffolding, (4) misassembly detection, and (5) conservative scaffolding. The revised A5 pipeline uses a new version of idba_ud that uses read pairing information, and that has been modified to accept reads up to 500 nt long and to construct *de Bruijn* graphs with *k*-mers up to 500 nt. These modifications provide substantial improvements in assembly contiguity.

The genome was annotated with the RAST annotation system using FigFAM release 70[11]. Putative antibiotic resistance genes and other genes of interest identified by RAST annotation were manually curated using the NCBI ORF finder and iterative BLASTn and BLASTp searches.

Quality assurance

The A5 pipeline includes a quality checking step that detects putative misassemblies by identifying clusters of read pairs that map to disjoint locations in the assembled genome. This method did not detect any putative misassemblies.

Initial findings

Sequencing generated 1,702,236 read pairs for a total of 483,658,987 nt that were assembled to reconstruct the 5,139,229 bp genome of *E. coli* 2009-46 in 89 scaffolds, with a scaffold N50 of 161 kbp and an N90 of 30.8 kbp. The raw (unfiltered) coverage is 94x, and after read filtering the assembly has a median depth of coverage of 61x. The annotation of this assembly identified 5084 predicted CDS and 106 predicted RNA genes. 19 genes were identified as possibly missing from the assembly by the RAST system. The overall functional profile of the genome is shown in Figure 1. We conducted a phylogenetic analysis of *E. coli* 2009-46 using the PhyloSift

software[12] to identify the most closely related organism with an available reference genome. PhyloSift works by identifying homologs in the draft genome to a set of 37 genes that are universally conserved among bacteria and archaea and present in single copy. It then adds any homologs found in the draft genome to an existing multiple sequence alignment containing the 37 genes from a subset of all genomes publicly available in the NCBI and EBI databases that is chosen to span the phylogenetic diversity of these databases. The PhyloSift reference database includes only a single representative from groups of closely related organisms. To gain additional resolution in the *Escherichia*, we used PhyloSift to construct a multiple alignment of the 37 marker genes from all finished *E. coli* genomes available in the NCBI database as of September 2013. We then inferred a phylogeny from that alignment using FastTree2[13]. The resulting analysis, shown in Figure 2, identified *E. coli* 536 as the most closely related isolate with a finished genome available, although there was some uncertainty in the 37 gene alignment as to whether *E. coli* 2009-46 diverged on the same lineage as *E. coli* 536. We used the closely related genome of *E. coli* 536 as a reference for further comparative analysis.

The scaffolds of *E. coli* 2009-46 were reordered to match the order in the finished genome of the closely related strain *E. coli* 536 using the Mauve Contig Mover[14]. After reordering, the genomes had 82 predicted rearrangement breakpoints. Many of these cluster in regions containing annotated transposase genes and multi-copy transporter gene families, suggesting either homology-mediated rearrangement or misassembly has occurred at these repetitive sequences. To further characterize the structure of the genome we used the CGview webserver[15] to plot matches to annotated proteins and the GC skew of the genome, with scaffolds ordered according to the *E. coli* 536 reference. The CGview plot is shown in Figure 3. Of note, the GC skew in *E. coli* 2009-46 genome appears to fluctuate frequently. This pattern is in sharp contrast to the GC skew of the *E. coli* 536 reference, which shows a strong pattern coinciding with the chromosome's replication arms (data not shown). This suggests that either *E. coli* 2009-46 has undergone substantial genome rearrangement in the recent past, that the true genome arrangement may not match the *E. coli* 536 reference very closely, that undetectable misassembly errors exist in the *E. coli*2009-46 genome, or that some combination of these three situations exists. We note that our assembly pipeline contains a step to detect and fix misassembly errors; none were found in the genome of *E. coli* 2009-46.

Comparison of the gene content between *E. coli* 2009-46 and the finished *E. coli* 536 reference genome identified 164 annotated gene functions predicted to be present only in *E. coli* 2009-46. Included among these are several genes related to scavenging iron, a type VII secretion system, an IncF conjugation system, mediators of hyperadherence, and copper and mercury resistance genes. The full list of gene functions found only in 2009-46 and those which 2009-46 lacks relative to the reference isolate are listed in Additional files 1 and 2, respectively.

The *bla*$_{\text{TEM-1}}$ gene, conferring resistance to ampicillin, was present on scaffold 78.1 (2551 nt), while the *sul2-strA-strB* genes conferring resistance to sulphonamides and streptomycin was located on scaffold 67.1, which was 5064 nt long. Ends of both the scaffolds had a partial copy the insertion element IS*26*. The isolate also houses a clinical class 1 integron and two associated resistance genes on scaffold 71.1. One of the two resistance genes is a variant of dihydrofolate reductase (*dhfr*) gene which provides trimethoprim resistance to isolates and the other confers resistance to aminoglocoside antibiotics (*aadA*). However the scaffold, 71.1, is 3,863 nt long and also has a copy of IS*26* at both ends. We identified the presence of the 3'-CS of a class 1 integron on scaffold 58.1 (6679 nt long), that had an IS*26* on one end and an IS*1* element on the other. Presence of IS*26* elements at both ends of seven scaffolds has resulted in scaffold breaks around a region of the genome, which most likely harbours a complex resistance locus (CRL), during the assembly of the genome sequence. We were therefore unable to confirm the exact genomic location of the CRL or resistance genes.

Antibiotic resistance profile

The antibiotic resistance profile of *E. coli* 2009-46 was experimentally determined using the disk diffusion method. This strain was found to be resistant to Ampicillin, Trimethoprim, Sulphafurazole, Tetracycline, Streptomycin, Apramycin, Kanamycin, and Azithromycin. A full list of antibiotics tested and *E. coli* 2009-46 susceptibility is provided in Additional file 3.

To better understand the genomic basis for the observed antibiotic resistance traits, the genome was searched for specific genes known to confer antibiotic resistance. A listing of these genes and their presence or absence in *E. coli* 2009-46 is provided in Additional file 3.

Future directions

Improved efficiency of clinical genomics pipelines will eventually enable fine-scale epidemiological monitoring of *E. coli* outbreaks in real time. When fully developed, this capacity will influence clinical and public health decisions related to treatment and control of pathogen outbreaks. Genomic data such as is presented here will aid in the interpretation of data from future outbreaks.

Availability of supporting data

The draft genome assembly has been submitted to NCBI and is associated with BioSample accession SAMN02725027. Genome annotations are available from the RAST web server under accession 562.3620. The Illumina sequence reads have been deposited to the Short Read Archive under accessions SRX514806 and SRX514807. CDS: Coding DNA sequences; ORF: Open Reading frame; RAST: Rapid annotation using subsystem technology; A5: Andrew and Aaron's Awesome Assembly; gDNA: genomic DNA; nt: Nucleotides;

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JM extracted DNA and conducted all laboratory-based analysis on the isolate. PW and JS constructed Illumina libraries and sequenced them. AED and PRC conducted the assembly, analysis, and data deposition. AED, SD, JM and PRC wrote the paper. IGC and PRC provided general direction. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by a collaboration between the NSW Department of Primary Industries and the ithree institute. Strain was donated by the Natalie Miller of the Sydney Adventist Hospital Microbiology laboratory. We thank Fiona MacIver for providing comments on a draft of this manuscript.

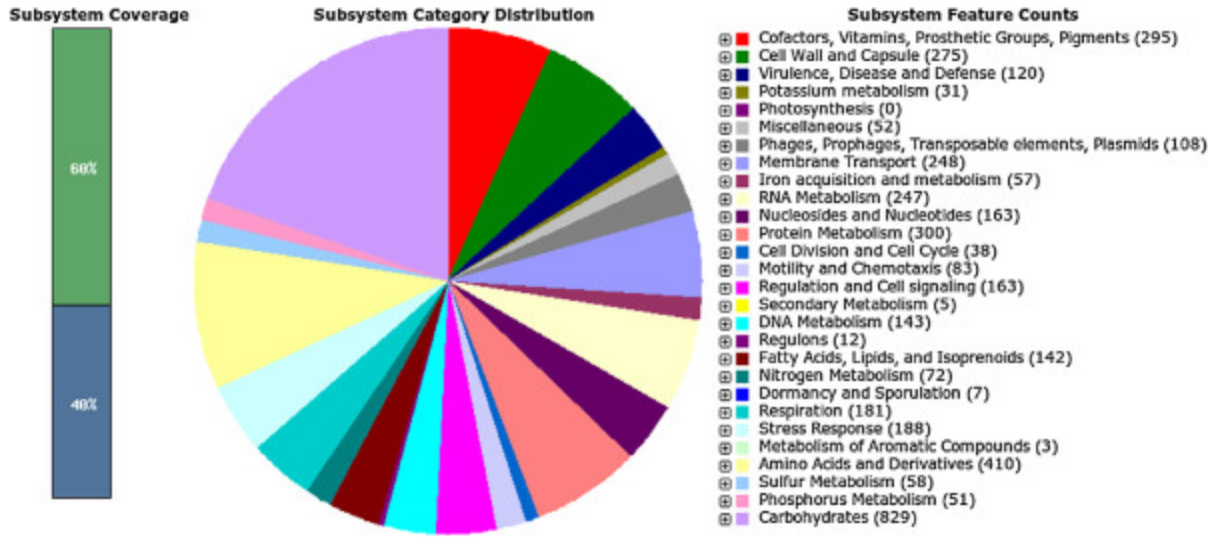


Figure 1: Subsystems in *E. coli* 2009-46.

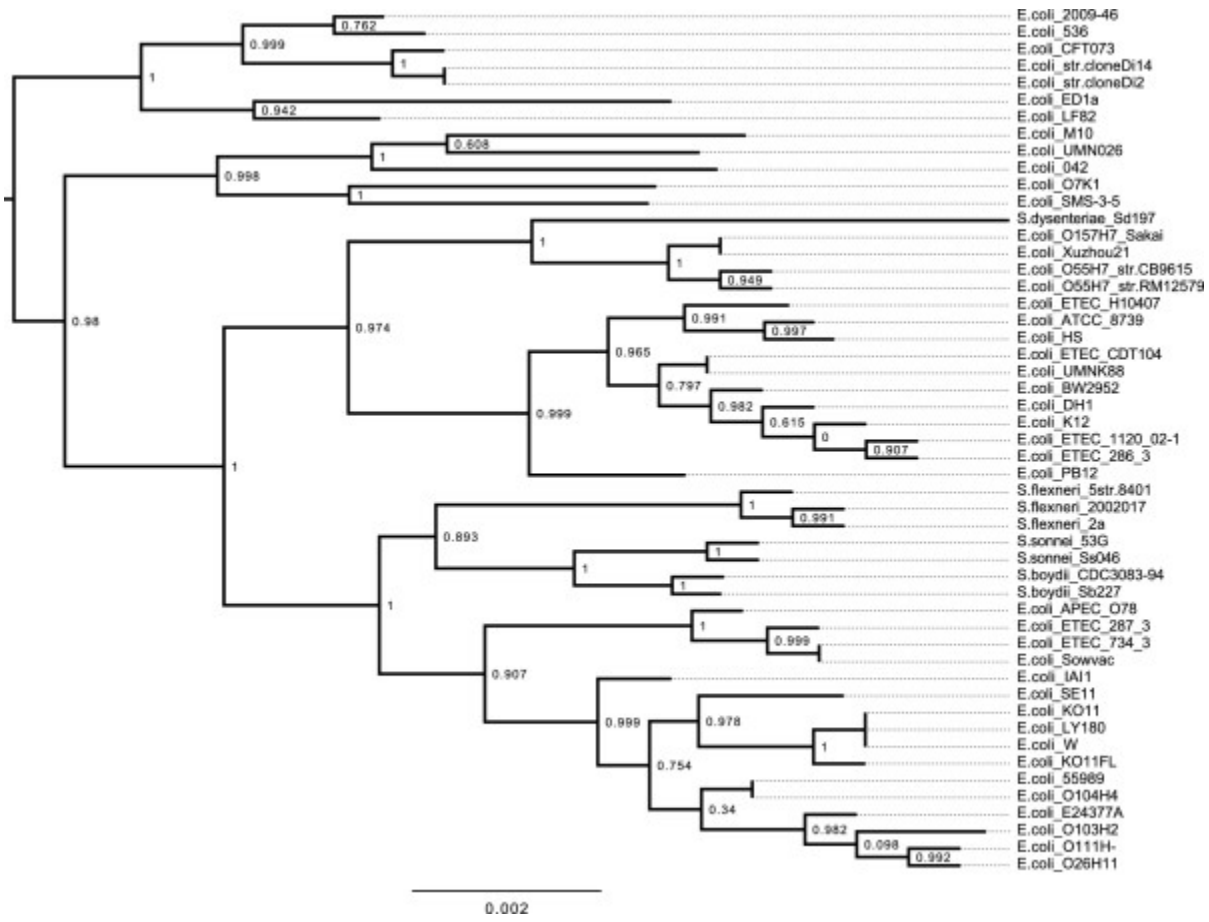


Figure 2: Phylogeny of ***E. coli*** and ***Shigella*** including the 2009-46 isolate. A phylogeny inferred on a concatenated set of codon alignments from 37 universally conserved genes is shown, as calculated by PhyloSift[12] and FastTree2[13]. The phylogeny has been rooted on the branch leading to *Salmonella* and internal nodes are labeled with SH-like support values.

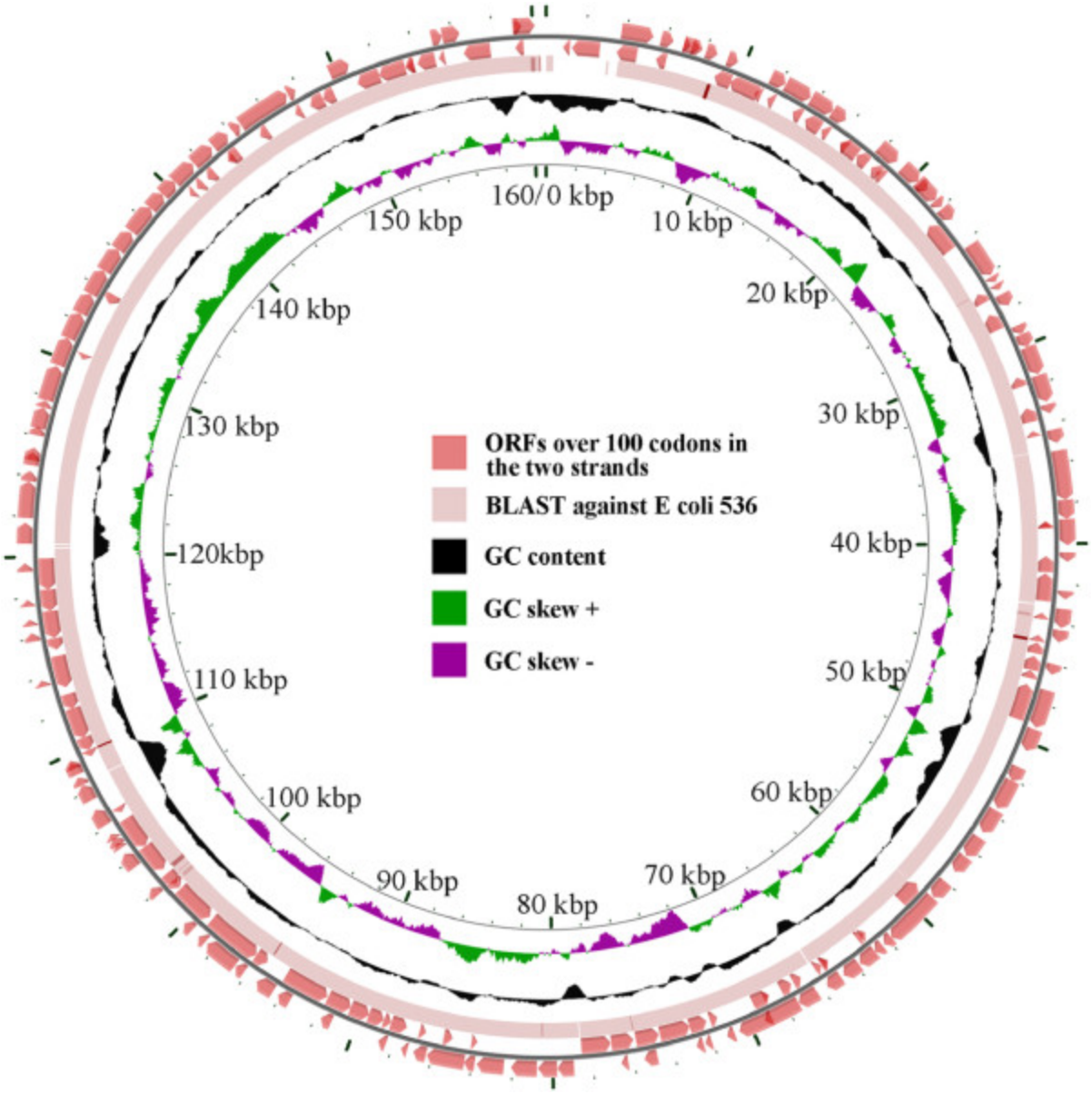


Figure 3: CGview plot of the ***E. coli*** 2009-46 genome. The two outermost circles in the figure contain a series of arrows in opposite directions representing predicted ORFs (greater than 100 codons) on the two strands of DNA sequence. The solid line, forming the third ring (from outside) indicates BLASTn analysis (set with a cutoff of 1e-10) of the isolate against the *E. coli* 536 genome. Relative GC content along the length of the genome is plotted as a graph in the black circle. The GC content clearly indicates multiple regions of GC content variation along the genome, possibly indicating lateral gene transfer events.

References

1. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect. 2003;5(5):449–456. doi: 10.1016/S1286-4579(03)00049-2.

2. Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J Infect Dis. 2000;181(5):1753–1754. doi: 10.1086/315418.

3. Johnson JR, Sannes MR, Croy C, Johnston B, Clabots C, Kuskowski MA, Bender J, Smith KE, Winokur PL, Belongia EA. Antimicrobial drug–resistant *Escherichia coli* from humans and poultry products, Minnesota and Wisconsin, 2002–2004. Emerg Infect Dis. 2007;13(6):838. doi: 10.3201/eid1306.061576.

4. Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L, Galanakis C, Reid-Smith RJ, Tellier P-P, Tellis PA, Ziebell K, Manges AR. Food reservoir for *Escherichia coli* causing urinary tract infections. Emerg Infect Dis. 2010;16(1):88. doi: 10.3201/eid1601.091118.

5. Jakobsen L, Garneau P, Bruant G, Harel J, Olsen S, Porsbo LJ, Hammerum A, Frimodt-Møller N. Is *Escherichia coli*urinary tract infection a zoonosis? Proof of direct link with production animals and meat. Eur J Clin Microbiol Infect Dis. 2012;31(6):1121–1129. doi: 10.1007/s10096-011-1417-5.

6. Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli*from Northwest England. J Antimicrob Chemother. 2012;67(2):346–356. doi: 10.1093/jac/dkr451.

7. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, Gilbert GL. Escherichia coli sequence type 131 as a prominent cause of antibiotic resistance among urinary escherichia coli isolates from reproductive-age women. J Clin Microbiol. 2013;51(10):3270–3276. doi: 10.1128/JCM.01315-13. doi:10.1128/JCM.01315-13.

8. Abraham S, Wong HS, Turnidge J, Johnson JR, Trott DJ. Carbapenemase-producing bacteria in companion animals: a public health concern on the horizon. J Antimicrob Chemother. 2014;69(5):1155–1157. doi: 10.1093/jac/dkt518.doi:10.1093/jac/dkt518.

9. Platell JL, Trott DJ, Johnson JR, Heisig P, Heisig A, Clabots CR, Johnston B, Cobbold RN. Prominence of an O75 clonal group (clonal complex 14) among non-ST131 fluoroquinolone-resistant *Escherichia coli* causing extraintestinal infections in humans and dogs in Australia. Antimicrob Agents Chemother. 2012;56(7):3898–3904. doi: 10.1128/AAC.06120-11.

10. Tritt A, Eisen JA, Facciotti MT, Darling AE. An integrated pipeline for *de novo* assembly of microbial genomes. PLoS ONE. 2012;7(9):42304. doi: 10.1371/journal.pone.0042304.doi:10.1371/journal.pone.0042304.

11. Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass E, Kubal M, Meyer F, Olsen G, Olson R, Osterman A, Overbeek R, McNeil L, Paarmann D, Paczian T, Parrello B, Pusch G, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008;9(1):75. doi: 10.1186/1471-2164-9-75.doi:10.1186/1471-2164-9-75.

12. Darling AE, Jospin G, Lowe E, Matsen IV FA, Bik HM, Eisen JA. Phylosift: phylogenetic analysis of genomes and metagenomes. PeerJ. 2014;2:243.

13. Price MN, Dehal PS, Arkin AP. Fasttree 2–approximately maximum-likelihood trees for large alignments. PloS one. 2010;5(3):9490. doi: 10.1371/journal.pone.0009490.

14. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics. 2009;25(16):2071–2073. doi: 10.1093/bioinformatics/btp356.

15. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36(suppl 2):181–184.

Chapter 3: Multidrug resistant uropathogenic *Escherichia coli* ST405 with a novel, composite IS*26* transposon in a unique chromosomal location

Multidrug resistant uropathogenic *Escherichia coli* ST405 with a novel, composite IS*26* transposon in a unique chromosomal location

Piklu Roy Chowdhury†,[1,2], Jessica McKinnon†[1], Michael Liu[1] and Steven P. Djordjevic[1]

[1]The ithree institute, University of Technology Sydney, NSW 2007, Australia. [2]NSW Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Menangle, Australia.

† These authors contributed equally to this study

Abstract

Escherichia coli ST405 is an emerging urosepsis pathogen, noted for carriage of $bla_{\text{CTX-M}}$, bla_{NDM}, and a repertoire of virulence genes comparable with O25b:H4-ST131. Extraintestinal and multidrug resistant *E. coli* ST405 are poorly studied in Australia. Here we determined the genome sequence of a uropathogenic, multiple drug resistant *E. coli* ST405 (strain 2009-27) from the mid-stream urine of a hospital patient in Sydney, Australia, using a combination of Illumina and SMRT sequencing. The genome of strain 2009-27 assembled into two unitigs; a chromosome comprising 5,287,472 bp and an IncB/O plasmid, pSDJ2009-27, of 89,176 bp. *In silico* and phenotypic analyses showed that strain 2009-27 is a serotype O102:H6, phylogroup D ST405 resistant to ampicillin, azithromycin, kanamycin, streptomycin, trimethoprim and sulphafurazole. The genes encoding resistance to these antibiotics reside within a novel, mobile IS*26*-flanked transposon, identified here as Tn*6242*, in the chromosomal gene *yjdA*. Tn*6242* comprises four modules that each carries resistance genes flanked by IS*26,* including a class 1 integron with *dfrA17* and *aadA5* gene cassettes, a variant of Tn*6029*, and *mphA*. We exploited unique genetic signatures located within Tn*6242* to identify strains of ST405 from Danish patients that also carry the transposon in the same chromosomal location. The acquisition of Tn*6242* into *yjdA* in ST405 is significant because it i) is vertically inheritable; ii) represents a reservoir of resistance genes that can transpose onto resident/circulating plasmids; and iii) is a site for the capture of further IS*26*-associated resistance gene cargo.

Keywords: IS*26*, Compound transposon, Class 1 integrons, Antibiotic resistance, *Escherichia coli* ST405, Tn*6242*.

Supplementary material available at https://www.frontiersin.org/articles/10.3389/fmicb.2018.03212/full#supplementary-material.

1. Introduction

Escherichia coli are commensals of vertebrate species but pathovars have evolved that cause both intestinal (intestinal pathogenic *E. coli*; IPEC) and extraintestinal infections in humans and in food producing and companion animals. Extraintestinal pathogenic *Escherichia coli* (ExPEC) have a faecal origin but have acquired the ability to colonise urinary tract epithelium causing cystitis, pyelonephritis and sepsis and are a leading cause of meningitis and wound infections[1, 2]. ExPEC are among the most frequently isolated Gram negative pathogens[3] and are increasingly resistant to broader classes of antimicrobials.

IS*26* plays a key role in the capture and spread of genes encoding resistance to different classes of antimicrobials, including drugs of last resort[4-15]. It can form independently mobile compound transposons containing tandem arrays of antimicrobial resistance genes[12, 16-18], often in association with class 1 integrons and diverse transposons, creating complex resistance gene loci (CRL). IS*26* can also delete and invert DNA, often modifying the conserved regions (5′- CS and 3′-CS) of class 1 integrons that are a target of PCR assays used to detect them and the resistance gene cargo they carry[12, 18-21]. These events often create novel signatures that can be exploited to track the CRL they reside in[12, 19]. Notably, IS*26* promotes: i) the formation of cointegrate plasmids from separate replicons that carry virulence and resistance genes[22]; ii) plasmid fitness by deleting regions of DNA that incur a fitness cost to the host[23]; and iii) the co-assembly and mobilisation of resistance genes efficiently in a *recA*$^-$ background by preferentially integrating in regions of genomes with a pre-existing copy of IS*26*[7].

ESBL-producing, multiple drug resistant ST405-D is an emerging ExPEC pathogen[24-26]. A phylogenetic study of the ST405 clonal lineage, and an analysis of mobile genetic elements that play an important role in the capture and spread of antimicrobial resistance in this lineage, has not been undertaken. It is not known if ST405 is a globally dispersed emerging clone or if disease is caused by sporadic episodes by genetically divergent ST405 strains. As part of a larger study of multiple drug resistant uropathogenic *E. coli* (UPEC) from Sydney hospitals, two strains were shown to carry a unique genetic signature in a class 1 integron created by the insertion of an IS*26* in the 3′-CS. We characterised their genome sequences using both short- and long-read genome sequencing technologies and determined both to be ST405, phylogroup D with serotype

O102:H6 (ST405-D-O102:H6). These events create opportunities for the further accumulation of antimicrobial resistance genes and are being targeted for sequence analysis to better understand how multiple drug resistance is emerging in Australia. Our analyses show that the strains carried a novel IS*26*-flanked composite transposon encoding resistance to multiple antibiotics in a unique chromosomal locus. Two ST405 genomes from Denmark with near identical structures were identified by interrogating *E. coli* genomes in the RefSeq database for molecular signatures unique to Tn*6242.*

2. Materials and Methods

2. 1 Strains, isolation and culture conditions

The *Escherichia coli* strains 2009-27 and 2009-30 were recovered one day apart from mid-stream urine of a patient being treated at the Sydney Adventist Hospital (SAN) in 2009. VITEK-2 species identification and phenotypic resistance profiling of the strains was performed at the SAN, testing for susceptibility to ampicillin, cephalexin, ciprofloxacin, gentamicin, nitrofurantoin and trimethoprim using established microbiological protocols. At UTS, the strains were additionally tested using the calibrated dichotomous sensitivity (CDS) method [27] against 15 antibiotics (ampicillin 25 µg, azithromycin 15 µg, cefotaxime 5 µg, cefoxitin 30 µg, chloramphenicol 30 µg, ciprofloxacin 5 µg, co-trimoxazole 25 µg, gentamicin 10 µg, imipenem 10 µg, nalidixic acid 30 µg, nitrofurantoin 200 µg, streptomycin 25 µg, sulphafurazole 300 µg, tetracycline 10 µg and trimethoprim 5 µg) and using *Escherichia coli* ACM 5185 as a reference control.

2.2 DNA purification

For PCR and Illumina sequencing, genomic DNA was extracted from 2 ml overnight cultures growing in Lysogeny Broth (LB) broth supplemented with 50 µg/ml ampicillin using an ISOLATE-II Genomic DNA kit (Bioline, Australia). Fosmid DNA and extrachromosomal DNA (IS*26* mediated DNA-loops) were extracted using the ISOLATE-II plasmid mini kit from 1.5 ml of LB culture supplemented with appropriate antibiotics. For Single Molecule Real Time (SMRT) sequencing genomic DNA from 1.8 ml of an overnight culture was isolated using a moBio genomic DNA kit (Qiagen, Germany).

2.3 Fosmid library construction, screening and PCR conditions

A fosmid library of strain 2009-27 was constructed using the CopyControl Fosmid Library Production Kit (Epicentre) with pCC2-Fos as the vector, following the kit protocol. The library was propagated in *recA*- EPI300 chemically competent *E. coli* cells supplied by Epicentre. Five hundred individual colonies were screened for the presence of *intI1* gene using (HS915 and HS916) primers listed in table S1. The single fosmid clone used in the inverse PCR/loop out assay was selected on the basis of end sequences of the *intI1* positive fosmid clones and targeted PCR cartograpgy experiments using primers listed in table S1.

PCR consisted of 10 µl of MyTaq red (Bioline) Master Mix, 0.5 µM of each primer and 0.2 µM of template DNA in DNAse-RNAse free water (final volume 20 µl). Cycling conditions comprised of: initial denaturation step at 94°C for 2 mins 30s, 30 cycles of denaturation (94°C for 30 s) annealing (60°C for 30 s) and extension (72°C for 5 min), with a final extension cycle of 72°C for 10 min. Annealing temperatures and extension step were varied depending on the melting temperatures of the primers and expected amplicon sizes. Table S1 lists the primers used. Figure S2 depicts the inverse PCR strategy used to identify circular elements comprising IS*26*-associated regions of the resistance locus.

2.4 Amplicon and whole genomes sequencing

Amplicons of interest were purified using a Promega Wizard® SV PCR and Gel Clean Up kit following the manufacturer's recommendations and sequenced using Sanger technology at the Australian Genome Research Facility, University of Queensland, in Brisbane. The genomes of strains 2009-27 and 2009-30 were initially sequenced at the ithree institute at UTS using a bench top Illumina MiSeq® sequencer and MiSeq V3 chemistry. The sequencing library was prepared following published protocols [28]. The 400 nt long paired end reads were assembled using an A5-MiSeq (ngopt_a5pipeline_linux-x64_20130919*)*.

DNA from strain 2009-27 was also sequenced using a PacBio RSII instrument at the Ramaciotti Centre for Genomics, UNSW, Sydney. Sequencing reads were assembled with HGAP and Quiver. The plasmid sequence was closed using Circlator[29] and polished with Illumina MiSeq raw read sequences using PILON[30]. Whole genome sequences of strain 2009-30 (Illumina) and strain 2009-27 (PacBio) are deposited in GenBank under accession numbers NXEQ00000000 and NXER00000000.

Preliminary genome annotations were generated using an online version of RAST[31] using FigFAM release 70. Putative virulence and antimicrobial resistance genes were identified using stand-alone BLASTn analyses and an in-house database, prior to manual verification using NCBI-ORF finder. Legitimate ORFs showed >95% sequence similarity (E-value 0.001) across 100% of the query sequence. Sequences of interest were characterised using iterative BLASTn and BLASTp searches[32].

2.5 Bioinformatics

Phylosift[33] was used to infer phylogentic relationships among ST405 strains 2009-27 and 2009-30 and 328 assembled *E. coli* ST405 genomes in Enterobase (http://enterobase.warwick.ac.uk/) as of 29-06-2017. A subset of 24 ST405 genomes that clustered together using marker gene phylogeny based PhyloSift analysis were subjected to core-genome alignment based phylogeny analysis using Parsnp[34] (https://github.com/marbl/Parsnp) and selecting the -c (all genomes) and -x recombination filter flags. The assembled genome of biosample SAMEA4559509 (sample ERS1458688) was downloaded from Enterobase and used as a reference. Comparative, whole genome analyses were performed using Mauve[35]. A SNP based phylogenetic tree of plasmid backbone sequences extracted from the longest locally collinear block derived from a progressiveMauve alignment was constructed using algorithms in FastTree_accu (http://darlinglab.org/blog/2015/03/23/not-so-fast-fasttree.html) and visualized using FigTree version 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Multilocus sequence typing was performed *in silico* (eMLST) using the PubMLST database (http://pubmlst.org/) and the Achtman *E. coli* MLST scheme (http://mlst.warwick.ac.uk/mlst/)[36]. Average nucleotide identity (ANI) was calculated using the ANI calculator portal at http://enve-omics.ce.gatech.edu/ani/index. SMRT sequences from strain 2009-27 were used or all comparative genomic analyses.

3. Results

3.1 Genomic analysis of strains 2009-27 and 2009-30

Illumina sequences of strains 2009-27 and 2009-30 assembled into 243 scaffolds (N50 = 89298; 47 × median coverage) and 191 scaffolds (N50 = 91427; 45 × median coverage) respectively.

SMRT sequences of strain 2009-27 (N50 = 20303) assembled into two unitigs; a chromosome comprising 5,287,472bp and a plasmid (pSDJ2009-27) of 89,176 bp. *In silico* analyses showed that strains 2009-27 and 2009-30 were ST405, serotype O102:H6 and phylogroup D (ST405-D-O102:H6). Both strains were phenotypically resistant to ampicillin, azithromycin, streptomycin, trimethoprim and sulphafurazole. ANI analysis of the two genomes indicated that strains 2009-27 and 2009-30 are indistinguishable from one another (99.9% mean identity and 100% median identity).

3.2 Plasmid analysis

pSDJ2009-27 typed as IncB/O/K/Z in plasmidFinder. The *repA* gene, encoding replication initiator protein, in pSDJ2009-27 was identical to *repA* in pR3521 (GU256641), a self-transmissible multi-drug resistant IncB/O plasmid encoding $bla_{\text{ACC-4}}$, $bla_{\text{SCO-1}}$, and $bla_{\text{TEM-1}}$ from a patient in Greece[37]. No antimicrobial resistance genes were found on pSDJ2009-27 but it contained IncI1 plasmid-related *tra*-genes essential for conjugative transfer. Notably, the sequence of *repA* was also identical to *repA* in pO26-Vir (FJ386569.1) and showed ≥99% similarity to *repA* sequences in plasmids pR3521 (GU256641), pEC3I (KU932021), pHUSEC411-like (HG428756) and pHUSEC41-1 (HE603110) associated with IPEC. BLAST analysis showed that pSDJ2009-27 shared ≥99% identity over 89% of the query sequence with pO26-Vir (FJ386569.1) from Shiga-toxin producing O26:H11 strain H30[38] (Figure S3). A progressiveMauve alignment (Figure S4) of the plasmid sequences with near identical *repA* genes revealed a ~ 60 kb conserved fragment of high sequence similarity. Phylogenetic analysis of the aligned regions of these plasmids (Figure S5) confirmed that pO26-Vir was the most closely related plasmid and originated from a recent common ancestor. A comparative bi-directional peptide BLAST analysis of all ORFs identified by RAST within the 60 kb fragment in pSDJ2009-27 with those on pO26-Vir, pEC3I, pHUSEC411-like and pHUSEC41-1 indicated that they contain genes for IncI plasmid maintenance, stability and transfer (Table S6). pO26-Vir is a large (168-kb) mosaic plasmid that carries putative EHEC virulence genes *toxB*, *katP*, *ehxA* and *espP*. Notably, ~ 45-kb of pO26-Vir containing these genes is absent in pSDJ2009-27 (Figure S3 and Table S7). We have recently described pO26-CRL[39] (represented as the inner green circle in figure S3) from an O26:H- EHEC isolated from a patient with haemorrhagic colitis and pO26-CRL-125 (Venturini et al., 2013), that were also similar to pO26-Vir (Figure S3).

3.3 Characterisation of the CRL in strain 2009-27

The class 1 integron in strains 2009-27 and 2009-30 is flanked by direct copies of IS*26,* one of which was located 209 nucleotides downstream from the start site of the *sul1* gene in the 3`-CS, rendering *sul1* inactive. The variable region of the integron carried *dfrA17* and *aadA5* genes encoding resistance to trimethoprim and streptomycin. A PCR developed previously to identify novel IS*26*-mediated genetic signatures within the 3′-CS of class 1 integrons using a primer in *attI1* (L1 primer) in the 5′-CS and a primer in IS*26 tnpA* (JL-D2)[12, 19] generated a unique amplicon of 2425-bp and this was exploited to track this unique genetic locus.

Analysis of the SMRT sequences from strain 2009-27 indicated it had a chromosomally located CRL, 19,774-bp (nucleotides 1732675 to 1752448, unitig_0 (NXER00000000) in length. The CRL is flanked by direct copies of IS*26* and located within *yjdA* in the chromosome, a site not known for insertion of laterally acquired DNA (Figure 1). An eight nucleotide direct duplication (TCTCACAG) was identified (Figure 1) flanking the CRL indicating that IS*26* was responsible for its movement. The IS*26*-flanked transposon is registered as Tn*6242* in the transposon database (https://transposon.lstmed.ac.uk/tn-registry).

The structure of the CRL is modular, comprising five copies of IS*26* that flank modules 1-4 containing combinations of different antimicrobial resistance genes. Module 1 comprises *repA-repC-sul2-strA-strB* genes bounded by two inwardly orientated IS*26* elements, a structure that forms part of the Tn*6029* family of transposons[12, 19, 40, 41]. Module 2 is 6924-bp and comprises Δ*sul1*, Δ*orf5* interrupted by the inverted repeat of Tn*501* (IR_{Tn501}), *chrA*, *orf98*, the IR*t* inverted repeat typical of clinical class 1 integrons, IS*6100,* and the macrolide resistance operon *mphR-mrx-mphA*. The sequence between *Δsul1* and the macrolide resistance operon is frequently reported in association with class 1 integrons on different *E. coli* plasmids from geographically distant sources (GenBank accession numbers LT632321.1; FQ482074.1; EU935739.1). Beyond the macrolide resistance operon is a third 1402-bp Tn*2*-derived IS*26* flanked module (module 3) comprising $bla_{TEM\text{-}1b}$ and a ΔTn*2* transposase gene. $bla_{TEM\text{-}1b}$-associated Tn*2*-derivatives typically form part of the Tn*6029* family of transposons[12], although this 1402-bp fragment is unique (see later). A class 1 integron with a truncated copy of *intI1* comprises module 4 in the CRL (Figure 1). Δ*intI1* comprises 657 nucleotides and the deletion event presumably was associated with insertion of IS*26*. The recombination functional of Δ*intI1* is non-functional in strains 2009-27 and

2009-30 and the integron is not expected to integrate or exchange gene cassettes. The P_c promoter remains intact however and is expected to induce the expression of *aadA5* and *dfrA17* resistance genes, an observation consistent with phenotypic resistance to trimethoprim and streptomycin.

3.4 Prevalence of Tn*6242*

A 24.1-kb DNA fragment comprising 19.8-kb of the CRL and about 5 kb flanking both ends of *yjdA* gene was used to interrogate the GenBank RefSeq database. Two *E. coli* genomes (accession numbers NZ_KE701406.1 and NZ_KE699379.1) showed ≥ 99% sequence identity across 100% of the input query sequence (Figure S8). NZ_KE701406.1 and NZ_KE699379.1 were ST405-D-O102:H6 and were from urine and blood of unrelated Danish patients respectively. Each strain carried the four resistance gene modules of Tn*6242*, including the class 1 integron and associated cassette array, on one supercontig. Barring small deletions that may be products of assembly errors, these sequences were indistinguishable (Figure S8).

3.5 Comparative phylogenomics of ST405 strains

To identify a cohort of ST405 strains related to strains 2009-27 and 2009-30 we initially performed a marker gene-based phylogeny analysis of 328 assembled *E. coli*-ST405 genomes from Enterobase (26th July, 2017) using Phylosift[33]. Strains 2009-27 and 2009-30 clustered with 24 ST405 strains (Figure S9) from human infections from unrelated geographic locations. Core genome alignment based phylogeny analysis of these 24 ST405 strains (Figure 2) identified e01776 to be the closest relative of 2009-27 (1414 SNP differences) while four strains, from patients in the USA with blood infections, showed between 1662 and 1665 SNP differences. Fifteen of the 24 strains (Table S10) had an identical class 1 integron and the remaining nine had *intI1* interrupted by IS*26.* The insertion site of IS*26* in *intI1* occurred at two separate locations both of which were different to Δ*intI1* in Tn*6242* (Table S10).

A BLASTn search (≥ 95% identity over 100% of the query sequence) of antimicrobial resistance and virulence genes in the genome of 2009-27 against an in-house database of resistance genes identified $bla_{\text{CTX-M88}}$ and $bla_{\text{OXA-1}}$ in the five ST405 genomes that clustered with 2009-27 (Table 1). Several of the five *E. coli* ST405 genomes carry all the resistance genes characteristic of Tn*6242* (*dfrA17*, *aadA5*, *sul1*, *mph(A)*, *sul2*, *strA* and *strB*) on the same scaffold. Table S11 depicts an alignment of the 19.7-kb region spanning Tn*6242* with the seven closely related *E. coli* ST405 genomes identified here. The four blood-borne *E. coli* ST405 strains had a copy of

the class 1 integron with a Δ*intI1* variant, as well as the IS*6100*-associated macrolide resistance module that abuts the 3`-CS with *orf5*, *chrA* and *orf98* genes on the same scaffold. Strain eo1776 did not carry the macrolide resistance module. Notably, all five strains contained an intact copy of *yjdA* and had a variable repertoire of virulence genes (Table S12). Comparative plasmid profiling of the five closely related genomes with strain 2009-27 suggested that *E. coli* ST405 strains 09-90543 and 09-90544 also had an IncB/O/K/Z plasmid replicon in addition to IncFII and pCol replicons, while blood isolates 09-0464 and blood 09-0292 carried IncFII and pCol replicons. Strain eo1776 was unusual in that it only carried an IncFII replicon. These data suggest that the class 1 integron may be associated with IncFII plasmids in this collection.

3.6 Evolution of Tn*6242*

In module 4 of Tn*6242* the sulphonamide resistance gene *sul1* is interrupted by a fragment of DNA encoding $bla_{\text{TEM-1B}}$ and the macrolide resistance operon in module 1 is flanked by oppositely facing copies of IS*26*. The reverse complement of the last eight bases of the *sul1* gene (TCTAAGAG) flanks one arm of module 2 while the complimentary sequence CTCTTAGA flanks the other end (Figure 1) of module 3. These observations, coupled with the fact that the orientation of module 2 and 3 is inverted in relation what is often seen in typical clinical class 1 integrons, suggest that modules 2 and 3 have undergone an IS*26*-mediated inversion, an event that can occur in regions of DNA flanked by oppositely-facing copies of IS*26*[12]. To explain the genealogy of Tn*6242* (Figure 3), a hypothetical progenitor of 10,698 bp, comprising an inverted copy of modules 2 and 3 that abuts the 3′-CS of the class 1 integron with *sul1* intact, was assembled *in-silico*. We also removed a copy of IS*26* that presumably inserted into the 3′-end of *sul1* in the progenitor. A BLASTn analysis was undertaken to determine if our hypothetical progenitor has been described previously. The analysis identified matches (100%) to integrons in plasmids ‘p’ from strain NCTC 13441 (LT632321.1), pETN48 from *E. coli* ST405-O102 (FQ482074.1), pEK499 (EU935739.1) from ST131-O25:H4, and pEC958 from *E. coli* ST131-O25b:H4 (HG941719.1) from Australia. From the progenitor sequence there are several features of Tn*6242* that indicate it may have formed in one of two ways. Option 1 in Figure 3 depicts a copy of IS*26* inserting into *sul1*, generating a target site 8-bp duplication (TCTAAGAG), followed by the insertion of a Tn*6029*-family transposon[7, 42]. In option 2 (Figure 3), a Tn*6029*-family transposon, rather than IS*26,* inserted in the *sul1* generating the same 8-bp target site duplication. The resulting structure, identified as the evolutionary intermediate in Figure 3,

positions 209 nucleotides at the 3′-end of *sul1* gene away from the rest of the *sul1* gene. Tn*6029*- and Tn*6026*-family transposons are widely dispersed within *Salmonella enterica* and *E. coli* populations in Australia and are known to be associated with plasmid-encoded, mercury resistance transposons[18, 19]. Strain 2009-27 does not contain a mercury resistance transposon, indicating that the formation and mobilisation of the CRL was driven by IS*26*. In the evolutionary intermediate, two inwardly-facing copies of IS*26* flank *repA-repC-sul2-strA-strB* and the *mphR-mrx-mphA* module beyond with the intervening sequence (option 2, Figure 3), enabling us to trace an IS*26*-mediated inversion event, generating the structure seen in Tn*6242*.

3.7 Generation of laterally mobile translocatable units from Tn*6242*

It is conceivable that Tn*6242* can generate multiple translocatable units (TUs), each mobilising a different combination of resistance gene module flanked by direct copies of IS*26*. To investigate this, we cloned the CRL into a pCC2Fos fosmid vector and propagated the fosmid in a *recA*- background (*E. coli* strain EP1300). An inverse PCR strategy (Figure S2) was used to examine the ability of the module containing the macrolide resistance gene cluster and the integron-containing module to be rescued as a translocatable unit (TU). Sanger sequencing of inverse PCR amplicons that span across the IS*26* junctions indicated that all three TUs were generated (Figure S13). Therefore, pSDJ2009-27 could act as a vector for Tn*6242* and generate different combinations of laterally mobile resistance gene modules flanked by direct copies of IS*26* from Tn*6242*.

4. Discussion

Genome sequence analyses of ST405-D-O102:H6 strains 2009-27 and 2009-30 from the urine of the same patient that were resistant to ampicillin, azithromycin, streptomycin, trimethoprim and sulphafurazole is presented here. In these Sydney strains, we describe the structure of a novel IS*26*-derived and laterally mobile compound transposon, identified here as Tn*6242*. Tn*6242* comprises a class 1 integron, a variant of Tn*6029,* and an *mphA* operon and these are sufficient to encode resistance to a panel of antibiotics described above. Notably, Tn*6242* is located within a unique chromosomal location in *yjdA*. We predict a series of genetic events that led to the formation of Tn*6242* and showed how the genetic signatures formed during its evolution were exploited to identify two MDR strains of ST405 from Denmark that carry minor variants of Tn*6242* in *yjdA*.

Our data suggest that this lineage of MDR ST405-D-O102:H6 may be globally disseminated. Analyses of phylogenetically related ST405 genomes in Enterobase indicate that they carry CRL with similar genetic cargo but were not associated with *yjdA*. Since most IS*26*-associated resistance regions are plasmid-encoded, the evolutionary events that created Tn*6242* in *yjdA* are likely to be recent. It will be important to monitor IS*26*-flanked CRL that associate with *yjdA* as it may represent a novel and emerging hotspot for the insertion of other IS*26*-flanked elements in the chromosome of clinically important Enterobacteriaceae.

The *repA-repC-sul2-strA-strB* module in Tn*6242* is flanked by two inwardly facing IS*26* elements and shares 99% sequence identity with structures found in pSRC26 (GQ150541), pSRC27-H (HQ840942.1)[16], pO26-CRL (GQ259888.1), pO26-CRL-125 (KC340960.1), pO111-CRL-115 (KC340959.1)[18], all reported in Australia. However, the Tn*2* transposon in the $bla_{\text{TEM-1b}}$ module is 1402-bp and is much smaller than Tn*2* variants in Tn*6029* (1916-bp), Tn*6029B* (1831-bp) or Tn*6029C* (1820-bp)[12]. BLAST analysis confirmed that the 1402-bp variant is novel suggesting that the deletion likely arose as a consequence of the inversion event described above. Notably, sequences that aligned with ≥ 99.9% identity to the input query were plasmids from Australia that carry derivatives of Tn*6029B*. Collectively, our data suggests that Tn*6242* may be a locally derived IS*26*-flanked transposon.

Finally, the backbone of a resident IncB/O/K/Z plasmid (pSDJ2009-27) showed ≥99% sequence identity with pO26Vir from an O26:H11 Shigatoxigenic *E. coli* and pO26-CRL from an enterohaemorrhagic *E. coli* O26: H^- strain. STEC O26:H11/H- has been associated with Shiga toxin production and foodborne illness and are the second most frequent cause of EHEC-associated food-borne illness globally[43]. pO26-Vir and pSDJ2009-27 carry a type IV pilus biosynthesis locus (*pil*) comprising *pilL* - *pilV* that may be important for biofilm formation[44]. ExPEC have a faecal origin and plasmids circulating in IPEC may adapt to play important roles in ExPEC virulence. To our knowledge, such an observation has not been reported and this is the first time an IPEC virulence plasmid has been identified in uropathogenic *E. coli*.

Acknowledgements

We wish to acknowledge Natalie Miller and Nola Hitchik at the SAN Pathology Laboratories for supplying strains 2009-27 and 2009-30 and associated metadata. We wish to thank Max Cummins for assistance with the standalone BLASTn analysis presented in the phylogenomics section of this manuscript. PacBio sequencing was undertaken at the Ramaciotti Centre for Genomics at the University of New South Wales. 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 ithree institute at the University of Technology Sydney. JM is the recipient of the Australian Postgraduate Research Award. The PacBio RSII sequencer used in this study was secured under ARC project LE150100031.

Author contributions

PRC and SPD conceptualised and designed the study and co-wrote the manuscript. PRC and JM analysed the data. JM also assisted in the generation of the first draft of the manuscript. ML was involved in the generation of short and long read sequencing of the genomes.

Conflict of Interest

None to declare.

Ethics statement

This study does not require any ethics approval. The isolates in the study came from a larger collection of E. coli isolates set aside by the hospital microbiology laboratory for our research. They were collected in accordance with Hospital policy as part of routine microbiology of hospital patients. We received de-identified isolates so patient identity is not known. There was no requirement for ethics approval on de-identified samples when they were collected.

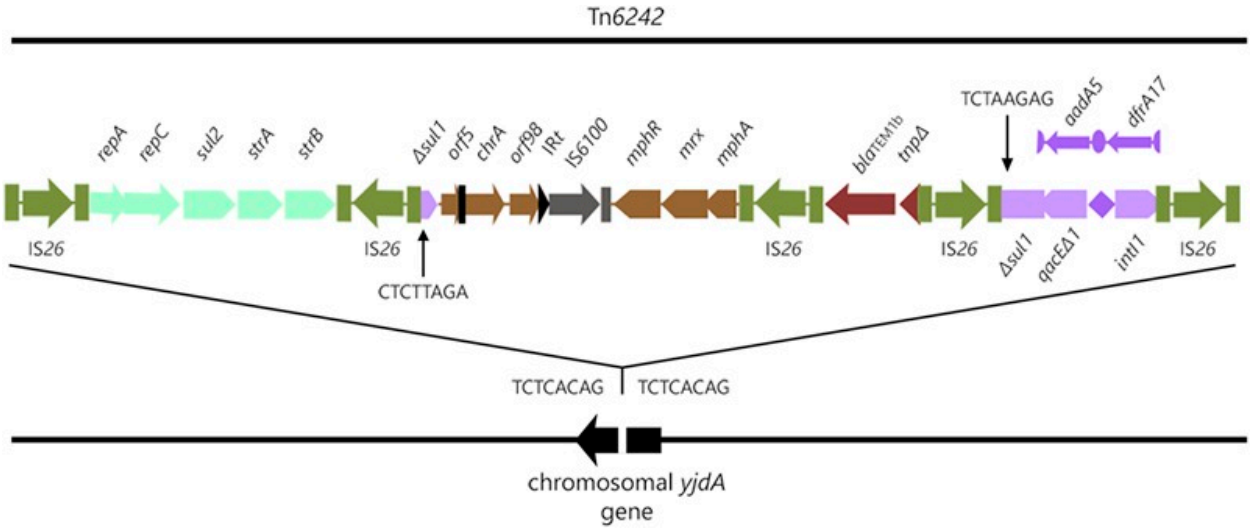


Figure 1: Structure of Tn*6242* in the chromosome of ST405 strain 2009-27.

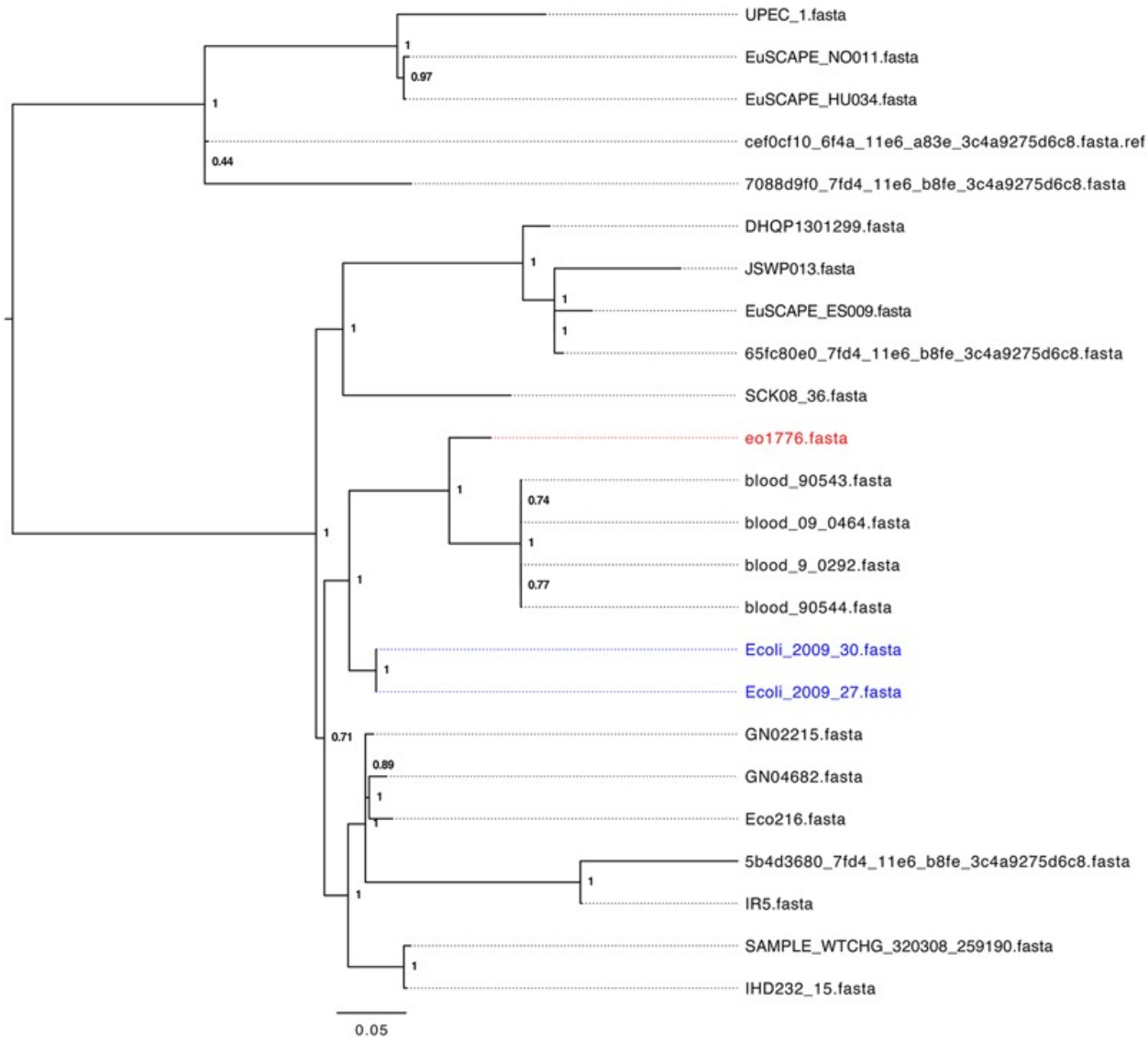


Figure 2: Parsnp tree analysis of *E. coli* ST405 genomes phylogenetically related to strain 2009-27.

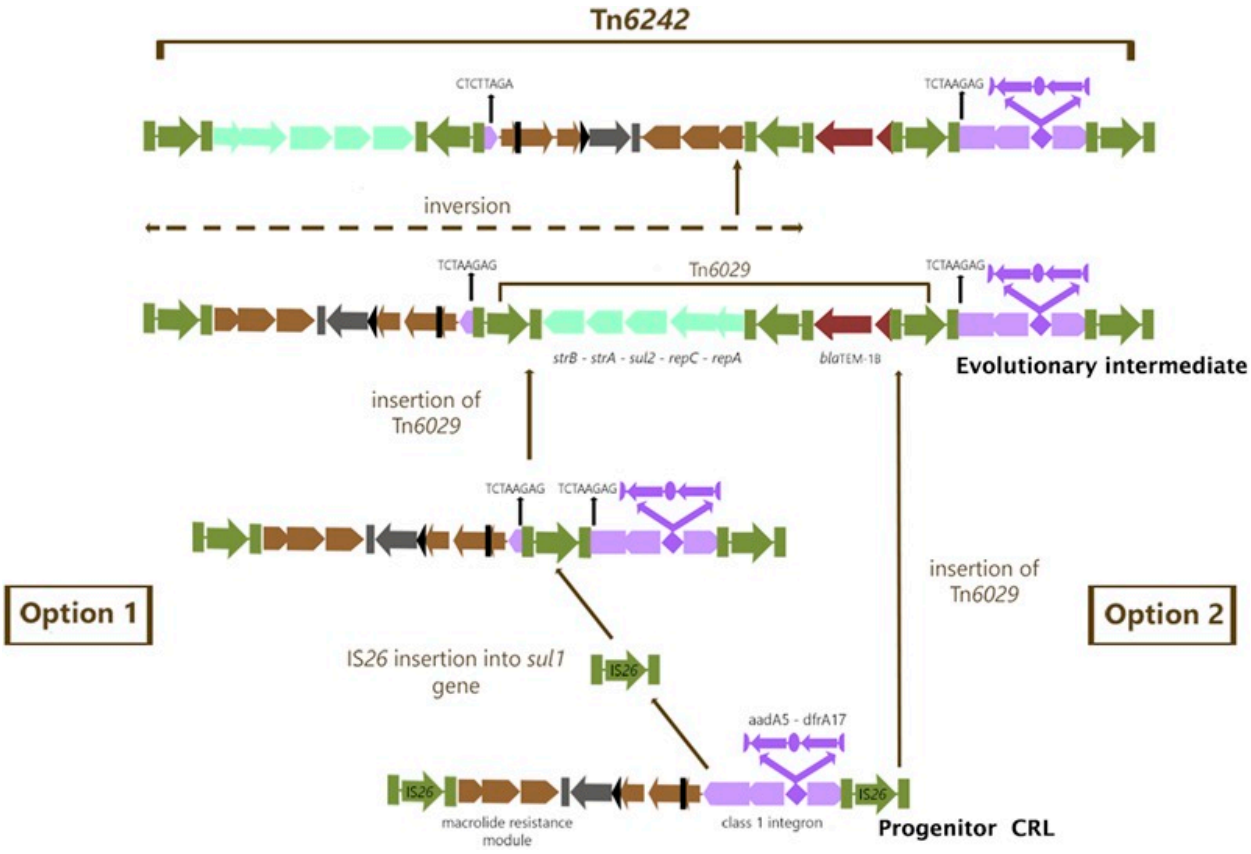


Figure 3: Genealogy of Tn*6242*. Green colour arrows bounded by green bars indicate IS*26*. The aqua cloured group of arrows indicate module 1 consisting of the *repA-repc-sul2-strA-strB* genes, the light brown coloured arrows indicate module 2 containing *orf5-chrA*-IS*6100-mphR-mrx-mphA* genes, the chocolate coloured gene indicates module 3 with the bla_{TEM1b} gene and the violet coloured arrows indicate the class 1 integron associated Δ*intI1-dfrA17-aadA5-qacE*Δ*1-sul1*Δ genes described as module 4. For more details of the genes and their orders please refer to figure 1.

Table 1: Resistance genes in ST405 genomes. Numbers indicate the percentage identity across 100% of the input query sequence. An asterisk (*) indicates the resistance genes present within the boundaries of Tn*6242*. △ indicates partial deletion at one end of the gene due to breaks in the scaffolds of the genome sequence.

Resistance genes	**2009-27**	**blood-09-0464**	**blood-9-0292**	**blood-90543**	**blood-90544**	**eo1776**
tet(B)(AF326777)	x	100	100	100	100	100
*tet(A)(*AJ517790)	x	100	100	100	100	x
sul2 (GQ421466) *	99.63	100	100	100	100	x
sul2 (HQ840942) *	100	99.63	99.63	99.63	99.63	x
strA (AF321551) *	100	100	100	100	100	100
strB (M963920) *	99.88	99.88	99.88	x	x	x
$bla_{\text{TEM-1A}}$* (HM749966)	99.65	x	x	99.77	99.65	x
$bla_{\text{OXA-1}}$ (J02967)	x	100	100	100	100	100
$bla_{\text{CTX-M-88}}$ (FJ873739)	x	99.89	99.89	99.89	99.89	99.89
aac(6')*Ib-cr* (DQ303918)	x	100	100	100	100	100
aacA4 (KM278199)	x	99.64	99.64	99.64	99.64	99.64
dfrA17 (FJ460238)*	100	100	100	100	100	100
aadA5 (AF137361)*	100	100	100	100	100	100
sul1 (CP002151)*	100△	100	100	100	100	100
mph(A) (D16251)*	100	100	100	100	100	x

References

1. Agyekum, A., Fajardo-Lubian, A., Ai, X., Ginn, A.N., Zong, Z., Guo, X., Turnidge, J., Partridge, S.R., and Iredell, J.R. (2016). Predictability of Phenotype in Relation to Common beta-Lactam Resistance Mechanisms in Escherichia coli and Klebsiella pneumoniae. *J Clin Microbiol* 54**,** 1243-1250.
2. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25**,** 3389-3402.
3. Bell, S.M., Pham, J.N., Rafferty, D.L., and Allerton, J.K. (2016). "Antibiotic susceptibility testing by the CDS method". (Australia: cdstest.net).
4. Beyrouthy, R., Robin, F., Lessene, A., Lacombat, I., Dortet, L., Naas, T., Ponties, V., and Bonnet, R. (2017). MCR-1 and OXA-48 In Vivo Acquisition in KPC-Producing Escherichia coli after Colistin Treatment. *Antimicrob Agents Chemother* 61.
5. Brolund, A., Edquist, P.J., Makitalo, B., Olsson-Liljequist, B., Soderblom, T., Wisell, K.T., and Giske, C.G. (2014). Epidemiology of extended-spectrum beta-lactamase-producing Escherichia coli in Sweden 2007-2011. *Clin Microbiol Infect* 20**,** O344-352.
6. Cain, A.K., and Hall, R.M. (2012a). Evolution of a multiple antibiotic resistance region in IncHI1 plasmids: reshaping resistance regions in situ. *J Antimicrob Chemother* 67**,** 2848-2853.
7. Cain, A.K., and Hall, R.M. (2012b). Evolution of IncHI2 plasmids via acquisition of transposons carrying antibiotic resistance determinants. *J Antimicrob Chemother* 67**,** 1121-1127.
8. Chen, C.M., Yu, W.L., Huang, M., Liu, J.J., Chen, I.C., Chen, H.F., and Wu, L.T. (2015). Characterization of IS26-composite transposons and multidrug resistance in conjugative plasmids from Enterobacter cloacae. *Microbiol Immunol* 59**,** 516-525.
9. Darling, A.E., Jospin, G., Lowe, E., Matsen, F.a.T., Bik, H.M., and Eisen, J.A. (2014a). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2**,** e243.
10. Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5**,** e11147.

11. Darling, A.E., Worden, P., Chapman, T.A., Roy Chowdhury, P., Charles, I.G., and Djordjevic, S.P. (2014b). The genome of Clostridium difficile 5.3. *Gut Pathog* 6**,** 4.
12. Dawes, F.E., Kuzevski, A., Bettelheim, K.A., Hornitzky, M.A., Djordjevic, S.P., and Walker, M.J. (2010). Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in Escherichia coli of human and animal origin. *PLoS One* 5**,** e12754.
13. Fratamico, P.M., Yan, X., Caprioli, A., Esposito, G., Needleman, D.S., Pepe, T., Tozzoli, R., Cortesi, M.L., and Morabito, S. (2011). The complete DNA sequence and analysis of the virulence plasmid and of five additional plasmids carried by Shiga toxin-producing Escherichia coli O26:H11 strain H30. *Int J Med Microbiol* 301**,** 192-203.
14. Gurnee, E.A., Ndao, I.M., Johnson, J.R., Johnston, B.D., Gonzalez, M.D., Burnham, C.A., Hall-Moore, C.M., Mcghee, J.E., Mellmann, A., Warner, B.B., and Tarr, P.I. (2015). Gut Colonization of Healthy Children and Their Mothers With Pathogenic Ciprofloxacin-Resistant Escherichia coli. *J Infect Dis* 212**,** 1862-1868.
15. Harmer, C.J., and Hall, R.M. (2015). IS26-Mediated Precise Excision of the IS26-aphA1a Translocatable Unit. *MBio* 6**,** e01866-01815.
16. Harmer, C.J., and Hall, R.M. (2016). IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* 1.
17. Harmer, C.J., Moran, R.A., and Hall, R.M. (2014). Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* 5**,** e01801-01814.
18. Hunt, M., Silva, N.D., Otto, T.D., Parkhill, J., Keane, J.A., and Harris, S.R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16**,** 294.
19. Johnson, J.R., Gajewski, A., Lesse, A.J., and Russo, T.A. (2003). Extraintestinal pathogenic Escherichia coli as a cause of invasive nonurinary infections. *J Clin Microbiol* 41**,** 5798-5802.
20. Jolley, K.A., and Maiden, M.C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11**,** 595.

21. Lajhar, S.A., Brownlie, J., and Barlow, R. (2017). Survival capabilities of Escherichia coli O26 isolated from cattle and clinical sources in Australia to disinfectants, acids and antimicrobials. *BMC Microbiol* 17**,** 47.
22. Liu, B.T., Song, F.J., Zou, M., Hao, Z.H., and Shan, H. (2017). Emergence of Colistin Resistance Gene mcr-1 in Cronobacter sakazakii Producing NDM-9 and in Escherichia coli from the Same Animal. *Antimicrob Agents Chemother* 61.
23. Mangat, C.S., Bekal, S., Irwin, R.J., and Mulvey, M.R. (2017). A Novel Hybrid Plasmid Carrying Multiple Antimicrobial Resistance and Virulence Genes in Salmonella enterica Serovar Dublin. *Antimicrob Agents Chemother* 61.
24. Matsumura, Y., Yamamoto, M., Nagao, M., Hotta, G., Matsushima, A., Ito, Y., Takakura, S., Ichiyama, S., and Kyoto-Shiga Clinical Microbiology Study, G. (2012). Emergence and spread of B2-ST131-O25b, B2-ST131-O16 and D-ST405 clonal groups among extended-spectrum-beta-lactamase-producing Escherichia coli in Japan. *J Antimicrob Chemother* 67**,** 2612-2620.
25. Matsumura, Y., Yamamoto, M., Nagao, M., Tanaka, M., Takakura, S., and Ichiyama, S. (2015). Detection of Escherichia coli sequence type 131 clonal group among extended-spectrum beta-lactamase-producing E. coli using VITEK MS Plus matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Microbiol Methods* 119**,** 7-9.
26. Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., and Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42**,** D206-214.
27. Papagiannitsis, C.C., Tzouvelekis, L.S., Kotsakis, S.D., Tzelepi, E., and Miriagou, V. (2011). Sequence of pR3521, an IncB plasmid from Escherichia coli encoding ACC-4, SCO-1, and TEM-1 beta-lactamases. *Antimicrob Agents Chemother* 55**,** 376-381.
28. Poolman, J.T., and Wacker, M. (2016). Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *J Infect Dis* 213**,** 6-13.
29. Porse, A., Schonning, K., Munck, C., and Sommer, M.O. (2016). Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Mol Biol Evol* 33**,** 2860-2873.

30. Reid, C.J., Roy Chowdhury, P., and Djordjevic, S.P. (2015). Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. *Plasmid* 80**,** 127-137.
31. Reid, C.J., Wyrsch, E.R., Roy Chowdhury, P., Zingali, T., Liu, M., Darling, A.E., Chapman, T.A., and Djordjevic, S.P. (2017). Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 3.
32. Riccobono, E., Di Pilato, V., Di Maggio, T., Revollo, C., Bartoloni, A., Pallecchi, L., and Rossolini, G.M. (2015). Characterization of IncI1 sequence type 71 epidemic plasmid lineage responsible for the recent dissemination of CTX-M-65 extended-spectrum beta-lactamase in the Bolivian Chaco region. *Antimicrob Agents Chemother* 59**,** 5340-5347.
33. Roy Chowdhury, P., Charles, I.G., and Djordjevic, S.P. (2015). 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* 10**,** e0115781.
34. Roy Chowdhury, P., Ingold, A., Vanegas, N., Martinez, E., Merlino, J., Merkier, A.K., Castro, M., Gonzalez Rocha, G., Borthagaray, G., Centron, D., Bello Toledo, H., Marquez, C.M., and Stokes, H.W. (2011). Dissemination of multiple drug resistance genes by class 1 integrons in Klebsiella pneumoniae isolates from four countries: a comparative study. *Antimicrob Agents Chemother* 55**,** 3140-3149.
35. Smith, J.L., Fratamico, P.M., and Gunther, N.W. (2007). Extraintestinal pathogenic Escherichia coli. *Foodborne Pathog Dis* 4**,** 134-163.
36. Srimanote, P., Paton, A.W., and Paton, J.C. (2002). Characterization of a novel type IV pilus locus encoded on the large plasmid of locus of enterocyte effacement-negative Shiga-toxigenic Escherichia coli strains that are virulent for humans. *Infect Immun* 70**,** 3094-3100.
37. 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.
38. Venturini, C., Beatson, S.A., Djordjevic, S.P., and Walker, M.J. (2010). Multiple Antibiotic Resistance Gene Recruitment onto the Enterohemorrhagic Escherichia coli Virulence Plasmid. *FASEB J.* 24**,** 1160-1166.

39. Venturini, C., Hassan, K.A., Roy Chowdhury, P., Paulsen, I.T., Walker, M.J., and Djordjevic, S.P. (2013). 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* 8**,** e78862.
40. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9**,** e112963.
41. Yau, S., Liu, X., Djordjevic, S.P., and Hall, R.M. (2010). RSF1010-like plasmids in Australian Salmonella enterica serovar Typhimurium and origin of their sul2-strA-strB antibiotic resistance gene cluster. *Microb Drug Resist* 16**,** 249-252.
42. Zhang, R., Sun, B., Wang, Y., Lei, L., Schwarz, S., and Wu, C. (2015). Characterization of a cfr-Carrying Plasmid from Porcine Escherichia coli That Closely Resembles Plasmid pEA3 from the Plant Pathogen Erwinia amylovora. *Antimicrob Agents Chemother* 60**,** 658-661.
43. Zong, Z., Ginn, A.N., Dobiasova, H., Iredell, J.R., and Partridge, S.R. (2015). Different IncI1 plasmids from Escherichia coli carry ISEcp1-blaCTX-M-15 associated with different Tn2-derived elements. *Plasmid* 80**,** 118-126.
44. Zurfluh, K., Poirel, L., Nordmann, P., Klumpp, J., and Stephan, R. (2015). First detection of Klebsiella variicola producing OXA-181 carbapenemase in fresh vegetable imported from Asia to Switzerland. *Antimicrob Resist Infect Control* 4**,** 38

Chapter 4: Whole genome analysis of ExPEC ST73 from a single hospital over a 2-year period identified different circulating clonal groups

Whole genome analysis of ExPEC ST73 from a single hospital over a 2-year period identified different circulating clonal groups

Bogema DR[1, 2]*, McKinnon J[2]*, Liu M[2], Hitchick N[3], Miller N[3], Venturini C[4], Iredell, J[4], Darling AE[2], Roy Chowdury P[2], Djordjevic SP[2]

[1] Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries, Menangle, NSW, 2568 Australia.

[2] The ithree Institute, University of Technology Sydney, NSW, 2007 Australia.

[3] San Pathology, Sydney Adventist Hospital, Wahroonga, NSW, 2076 Australia

[4] Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, The University of Sydney, Westmead, NSW, 2145 Australia

* authors contributed equally to this manuscript

Abstract

ST73 has emerged as one of the most frequently isolated extraintestinal pathogenic *E. coli* (ExPEC). To examine the localised diversity of ST73 clonal groups including their mobile genetic elements profile, we sequenced the genomes of 16 multiple drug resistant ST73 isolates from patients with urinary tract infection from a single hospital in Sydney, Australia between 2009 and 2011. Genome sequences were used to generate a SNP-based phylogenetic tree to determine the relationship of these isolates in a global context with ST73 sequences (n=210) from public databases. There was no evidence of a dominant outbreak strain of ST73 in patients from this hospital, rather we identified at least eight separate groups, several of which reoccur, over a two-year period. The inferred phylogeny of all ST73 strains (n=226) including the ST73 Clone D i2 reference genome shows high bootstrap support and clusters into four major groups which correlate with serotype. The Sydney ST73 strains carry a wide variety of virulence-associated genes but the presence of *iss*, *pic* and several iron acquisition operons was notable.

Keywords: Uropathogenic Escherichia coli, class 1 integron, Sydney ST73

All other information including data bibliography can be found at https://mgen.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000255#tab2.

Data Summary

1. All sequencing reads and assemblies for isolates sequenced in this study have been submitted to the ENA Sequence Read Archive and GenBank, respectively. GenBank, SRA accession numbers and URLs are included in Table S1.
2. Scripts used for the analysis of SNP phylogeny have been deposited in Github; (URL - https://github.com/bogemad/snp_phylogeny)
3. Figure S1 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477449 (URL – https://figshare.com/s/b720dcd41ca1c11fdae2)

4. Figure S2 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477485 (URL – https://figshare.com/s/0835e6174bf9f026ff61)
5. Table S1 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477461 (URL – https://figshare.com/s/e6ea61f9dde79e9b18dd)
6. Table S2 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477464 (URL – https://figshare.com/s/70de292aa98e806dd05f)
7. Table S3 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477473 (URL – https://figshare.com/s/b91bbc880eb31e5f90fc)
8. Table S4 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477476 (URL – https://figshare.com/s/3b8be84c33225bb32d78)
9. Table S5 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477479 (URL – https://figshare.com/s/7ba306327343fdeca167)

Impact Statement

ST73 is a major clonal lineage of ExPEC that causes urinary tract infections often with uroseptic sequelae but has not garnered substantial scientific interest as the globally disseminated ST131. Isolation of multiple antimicrobial resistant variants of ExPEC ST73 have increased in frequency, but little is known about the carriage of class 1 integrons in this sequence type and the plasmids that are likely to mobilise them. This pilot study examines the ST73 isolates within a single hospital in Sydney Australia and provides the first large-scale core-genome phylogenetic analysis of ST73 utilizing public sequence read datasets. We used this analysis to identify at least 8 sub-groups of ST73 within this single hospital. Mobile genetic elements associated with antibiotic resistance were less diverse and only three class 1 integron structures were identified, all sharing the same basic structure suggesting that the acquisition of drug resistance is a recent event. Genomic epidemiological studies are needed to further characterise established and emerging clonal populations of multiple drug resistant ExPEC to identify sources and aid outbreak investigations.

Introduction

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are phylogenetically diverse and comprise uropathogenic *E. coli* (UPEC), neonatal meningitis-causing *E. coli* (NMEC) and avian pathogenic *E. coli* (APEC). ExPEC account for ~75-95% of urinary tract infections (UTI). A proportion of these infections can spread from the urinary tract with invasion of epithelial cells in the bladder (cystitis) and kidney cells (pyelonephritis), and transmission to systemic circulation (blood sepsis), posing a serious threat to human health. ExPEC are enteric bacteria but their capacity to capture a wide array of virulence-associated genes (VAGs) by lateral gene transfer has expanded the repertoire of niches they colonise. ExPEC may carry diverse and often redundant combinations of VAGs whose impact on human health remains ill-defined. Epidemiological studies indicate that a subset of pathogenic *E. coli* lineages including ST73, ST131, ST405, ST393, ST69, ST95, ST10, ST38, and ST127[1-5] are responsible for most ExPEC infections[3, 6-8]. Carriage of combinations of virulence genes enhance virulence[9], however, carriage of antimicrobial resistance genes, particularly those encoding ESBLs and fluoroquinolones, as well as an ability to cause opportunistic infections in vulnerable (elderly) hosts, may also contribute to virulence. It is notable that none of these hypotheses have been experimentally validated[3, 6].

ExPEC have become the leading cause of blood sepsis in Europe[10]. Notable in this regard is the alarming rise in the incidence of ST73, now one of the most frequently isolated UPEC globally and the leading cause of bacteraemia in the East Midlands region of the UK[1, 11-14]. ST73 belongs to Clermont phylogroup B2 and is known to display different serogroups with serotype O6 predominating (ST73-O6-B2)[15, 16]. It has recently been suggested that the rise in the incidence of multiple drug resistant (MDR) ST73 in the UK is not due to the emergence of a dominant clone because they are genetically diverse and carry a different array of plasmids encoding resistance to multiple antimicrobials. Many recently described isolates of ST73 carry genes that encode extended spectrum β-lactamases and resistance to antimicrobials used in veterinary medicine[14]. This seems to be a recent adaptation in this ST as previously characterized ST73 isolates from cases of uncomplicated UTI sourced from Greece, Portugal, Sweden and the UK, resulted susceptible to most clinically relevant antimicrobials and most (75%) did not carry plasmids,

classic vehicles of MDR[17]. These data combined with the most recent findings seem to suggest that the rise in the carriage of MDR plasmids in ST73 may be a recent concerning event[14, 18].

Here we have characterised whole genome sequences of 15 class 1 integrase (*intI1*)-containing ST73 strains from a hospital in Sydney. To determine if these highly localised strains were from a limited number of clonal lineages, phylogenetic inferences were made by comparing SNP differences in core genomes shared by ST73 strains from our Sydney collection with those from six high-quality reference genome sequences and ST73 strains (n=204) from seven countries sourced from global sequence read archives. We also examined mobile and chromosomal genetic content within this localised isolate cohort to further examine their accessory genome diversity. We compiled the repertoire of antimicrobial genes and virulence genes and mapped the class 1 integron structures carried by these isolates. As carriage of the class 1 integrase is considered a reliable proxy for multiple drug resistance[19], we used S1-PFGE followed by Southern hybridization with an *intI1* probe to examine plasmid content and carriage of the class 1 integrase on plasmids.

Methods

Isolate source and culture conditions

Clinical samples in this project were from a larger collection obtained from Sydney Adventist Hospital from 2009-2011. Bacterial species were identified by the VITEK® 2 (bioMérieux) system at Sydney Adventist Hospital. For DNA extraction, strains were first grown on a Lysogeny Broth (LB) agar plate to isolate single colonies of which one was used to inoculate 2 mL of LB broth followed by shaking for 16 hours at 37°C. Antibiotic susceptibility testing for ampicillin, cefotaxime, chloramphenicol, streptomycin and sulfafurazole was performed via the CDS method[20]. These antibiotics were selected based on antibiotic resistance gene content inferred from genome sequencing.

Nucleic acid purification and whole genome sequencing

E. coli DNA was extracted using the Isolate II Genomic DNA Extraction kit (Bioline) according to the manufacturer's instructions. For each sample, tagmentation of genomic DNA, and PCR amplification of tagged DNA were performed in triplicate using the Nextera system (Illumina). Sequencing libraries were pooled, then cleaned and size selected using SPRI beads (Beckman Coulter). Normalization was guided by read counts obtained from a Nano flowcell run on a MiSeq instrument. Agilent 2100 Bioanalyzer, with High Sensitivity DNA kit was used to quantitate the pooled library before loading onto an Illumina HiSeq. Paired-end 150 bp reads were generated using the HiSeq 2500 v4 system.

Genome assembly and gene presence

Genome assembly was achieved with raw reads using the A5-miseq pipeline[21] and checked for consistency by additional assembly with SPAdes 3.9.0[22]. Antibiotic resistance and virulence-associated genes were identified from assembled genomes using BLASTn and SRST2[23]. Searches were performed against antibiotic resistance genes sourced from the ARG-ANNOT V3 database and a panel of virulence-associated genes identified from the Virulence Factors Database (VFDB) and literature searches[24, 25]. Serotyping was performed *in silico* with SRST2 using EcOH sequences supplied with the SRST2 package. Draft genome reads obtained from SRA were searched using SRST2 for a minimal set of marker genes derived from integron structures and commonly associated transposons characterised in the Sydney strains. Low-quality alignments based on SRST2 output were discounted (n=3).

Archived sequence read selection

All additional ST73 sequences not generated by this study were obtained from complete whole genome assemblies (n = 6)[26-30] and public sequence read archives (NCBI-EMBL). Raw Illumina reads sourced from ST73 isolates (n = 284) were considered for SNP-based phylogenetic analysis, including strains sequenced in this study from the Sydney Adventist Hospital (n = 16), isolates with host, source and isolation location meta-data identified from the EnteroBase database (n = 246; http://enterobase.warwick.ac.uk/; accessed 5/12/2016) and a previous ST73-focused study from the United Kingdom (n = 22)[11]. Samples were excluded if sequence type

could not be confirmed as ST73 using SRST2 (n = 4). Further samples were excluded (n = 30) if isolate status could not be confirmed by BioProject meta-data or where the description of methods could not be identified by an associated publication[31-38]. Samples were additionally excluded (n = 30) if they produced low reference genome coverage (>90%) in whole-genome alignments. Additional sample filtering of ST73 reads is described in Table S2.

S1-PFGE analysis

The complement of large (>20 kb) plasmids in each bacterial isolate was determined by S1 nuclease (Promega, Madison, WI, USA) digestion and pulsed-field gel electrophoresis (PFGE) as described previously[39, 40]. Southern blot hybridization was used to determine the genomic location of the *intI1* gene. PCR amplicons for *intI1* were obtained using published primers (int1F and int1R[41]) and labelled using the PCR DIG Probe Synthesis Kit (Roche, Mannheim, Germany). DNA was transferred from the S1-PFGE gel to a nylon membrane (GE Health, Little Chalfont, UK) using a VacuAid vacuum transfer apparatus (HybAid, Teddington, UK) and hybridisation was performed using the Roche DIG Filter Hybridization system (Roche, Mannheim, Germany) following manufacturer's instructions. Images were acquired on a ChemiDoc™ MP System (Bio-Rad Laboratories, Richmond, CA, USA).

SNP based phylogenetic analyses

Our initial attempts to examine ST73 phylogeny with our 16 genome sequences and 6 complete whole genome sequences (CFT073, ABU83972, ATCC25922, Nissle 1917, clone D i2 and clone D i14) using marker gene approaches[42] provided limited resolving power (Fig. S1). Consequently, to more appropriately examine these ST73 strains, we employed SNP-based phylogenetic methods.

For SNP-based phylogenetic trees, core genome alignments were generated with Snippy v3.1.0 (https://github.com/tseemann/snippy) using default options. Briefly, reads were mapped using bwa mem v0.7 to an ST73 reference genome. Raw alignments were processed with samtools v1.3.1 and variants called using freebayes v1.1.0. SNP derived genomes were reconstructed using vcftools v0.1.14, with low-coverage (< 10 x) and degenerate reference positions filtered. Recombinant regions were removed using Gubbins v2.20 (options -i 10)[43] yielding aligned, SNP-derived, recombination-filtered core genomes. From this alignment, core genome phylogenetic

trees were inferred by maximum-likelihood using RAxML v8.2.9[44]. Branch support was estimated by bootstrap analysis employing 100 replicate trees. Trees were rooted using the ultrametric tree method included with RAxML. Scripts used for the analysis of SNP phylogeny can be found online at https://github.com/bogemad/snp_phylogeny.

Initially we performed this analysis using only isolates sequenced in this study and a high-quality published ST73 reference (Fig. 1). To identify the most suitable reference genome for this purpose, we aligned reads individually to the six complete reference genomes (above) and examined reference sequence quality, core alignment lengths and final tree support values. Using this methodology, the most suitable reference was identified as clone D i2 which with archived public sequence reads generated a core genome of 3, 818, 344 bp, representing 75.8% of the ST73 clone D i2 sequence.

Results

Assembly information and statistics

The genome sequences of 16 ST73 strains from the Sydney Adventist Hospital were determined here. The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank and the sequence read archive. Assembly statistics as well as accession numbers, number of sequencing reads, and amount of sequencing data used to generate assemblies can be found in Table S1.

Public read high-throughput sequencing analysis

The isolates sourced from Sydney Adventist Hospital separated in several groups that closely aligned with *E. coli* serotype including O22:H1, O25:H1 and O6:H1. However, most of the Sydney isolates clustered within a larger O6:H1 group. To further interrogate observed diversity within the O6:H1 group of isolates and to place Sydney strains within a broader global context, we expanded the SNP-based phylogenetic analysis to include additional ST73 sequence reads obtained from public sequence read archives (NCBI-EMBL-DDBJ) and the six complete ExPEC genomes with the 16 Sydney Adventist Hospital genomes. The inferred phylogeny of the 226 strains (Fig. 2; see Fig. S2 for all strain labels and branch support values) shows strong major

branch support. Analysis of the SNP-derived phylogenetic tree shows correlation with observed serotypes O6:H1, O25:H1 and O22:H1, with most strains observed within the O6:H1 cluster. Strain 2011_82 could not be assigned an O-type from *in* silico serotyping but was identified as H1 and clustered most closely with O6:H1 isolates (Fig. 2 – Serotype).

Examination of ST73 phylogenetic structure reveals four significant clades (Fig. 2 and S2): one exclusively associated with serotype O6:H1 (Group A), one exclusively associated with serotype O2:H1 (Group B), one primarily serotype O6:H1 with a subclade of O18:H1 (Group C), and finally a polyphyletic group consisting of O2:H1, O6:H1, O22:H1, O25:H1 and O120:H31 serotypes (Group D). Strains sequenced from Sydney separated into eight distinct groups correlated with serotype (Fig. 2 – red labels). Isolates with O22:H1 and O25:H1 serotypes formed their own clusters, while serotype O6:H1 separated into 5 distinct clusters (O6-1: 2009_38/45; O6-2: 2009_6/2011_69; O6-3: 2009_44; O6-4: 2009_5/14/16/2010_64; O6-5:2009_8). Strain 2011_82 clustered with other O6:H1 strains and formed the final group (Ox).

Virulence profiles of Sydney strains

We identified differences in virulence gene profiles of *E. coli* strains examined in this study (Fig. 1). Significantly, we found that virulence gene profiles were largely consistent in strains from the same phylogenetic groups O25, O22, O6-1 – O6-5, Ox (Fig. 1, see Table S3 and S4 for greater detail). For adhesion-related genes, critical components of P fimbriae (*papACDEFGHJK*) were absent in six strains (Fig. 1). Additionally, in strains of group O25, genes encoding the Type I fimbriae major subunit (*fimA*), periplasmic chaperone (*fimC*), regulatory subunit (*fimE*) and the fimbriae-associated *fimI* were missing in blastn searches. Other genes encoding F1C fimbriae, Curli fibres, Type IV pili, *E. coli* common pili and the *fdeC* adhesion genes were present in all strains. The importance of FdeC as a putative virulence factor is underpinned by the observation that it is i) a broadly conserved, *E. coli* adhesin whose expression is upregulated on the surface of UPEC when it contacts host cells; and ii) a major target during humoral immune responses that significantly reduced kidney colonization in mice challenged transurethrally with UPEC strain 536 [45].

Iron acquisition is critical for the growth of ExPEC in low iron environments *in vivo* and it is not uncommon to identify genes linked to siderophore production and processing in UPEC. Complete

Enterobactin, Salmochelin, Yersiniabactin gene clusters were identified in all ST73 strains while Aerobactin genes were identified in all strains except those belonging to group O22. Genes for heme uptake including the *chu* operon and *hma* gene were present in all strains, as were those related to iron uptake such as the *sit* ABC transporter operon and ferric Yersinia uptake (*fyuA*) gene. In contrast, the putative iron uptake gene cluster *eitABCD* and adhesion/iron-uptake gene *ireA* were only identified in a subset of strains. In addition to iron uptake, genes encoding copper resistance have also been linked to virulence[46] and antimicrobial resistance[47]. The *cus* system, encoding a four-component copper efflux pump, was present and complete in all strains. However, in strain 2009_45, *cueR*, an important regulator controlling copper detoxification and efflux *copA* and *cueO* genes, was not located in all searches.

Larger differences were observed in the presence of toxin genes. Strains from group O25 and strain 2009_8 contained the highest number of toxin genes including cytotoxic necrotizing factor 1 (*cnf1*), the hemolysin (*hly*ABCD) cluster, hemolysin E (*hlyE*) and secreted autotransporter toxin (*sat*). Genes that have been previously shown to promote propagation of *E. coli* in blood, such as proteases *pic* and *tsh* and the increased serum survival (*iss*) gene were present in all strains as well as cellular invasion promoting *ibe* gene cluster. Closely related *hek* and *tia* genes, associated with epithelial cell invasion in Neonatal Meningitis-causing and enterotoxigenic *E. coli* respectively, are both found in separate strains. Furthermore, *tcpC* associated with immune modulation via inhibition of Toll/interleukin-1 (IL-1) receptor signalling, was only found in groups O6-1-5 and Ox.

Antibiotic resistance

All *intI1* positive isolates were tested for resistance to ampicillin, cefotaxime, chloramphenicol, streptomycin, sulfafurazole and trimethoprim using the CDS method. Strain 2011_82 did not have a class 1 integrase gene and was not tested. All strains were resistant to ampicillin, streptomycin and sulfafurazole (Table 1). Genes encoding resistance to these antibiotics were all accounted for in the genome sequence data by the class 1 integron-associated genes *aadA1* and *sul1* as well as one of three *bla* gene variants (Table S5). Only strain 2009_45 was resistant to the third-generation cephalosporin cefotaxime, likely due to the presence of the $bla_{\text{OXA-1}}$ gene. However, this resistance was not observed in strain 2009_38 which contained an almost identical antimicrobial resistance

region, suggesting this gene is not expressed in this strain. Interestingly, both of these strains also showed phenotypic resistance to chloramphenicol despite only 2009_45 containing a complete copy of the *catA1* gene. The full repertoire of antibiotic resistance genes found in the 16 Sydney ST73 strains is presented in Figure 1.

Structure of class 1 integrons in ST73 strains from Sydney

All locally sourced strains in this study, excepting 2011_82, were positive for a complete copy of the sulfonamide resistance gene *sul1*, a structural marker of the 3′-conserved segment (3′-CS) of class 1 integrons. Similarly, all strains contained the aminoglycoside resistance gene cassette *aadA1*. Strain 2011_82 was found to contain only a class 2 integron carrying the standard *dfrA1-sat2-aadA1* cassette array, resulting in trimethoprim resistance as tested by the hospital upon initial isolation.

There were three class 1 integron-containing resistance regions represented within our collection (Fig. 3), all containing the same base structures with minor variations. The first structure was identified in 11 out of 15 class 1 integron-containing isolates (Fig. 3A). It consisted of an In2 type class 1 integron with an *aadA1* gene cassette housed within an incomplete Tn*21* transposon, matching (99% sequence identity) to the sequence in the R100 plasmid identified in Japan in the 1950s from *Shigella flexneri* Accession NC_002134.1[48]. However, our structure bears an IS*26*-mediated partial deletion of the Tn*21 tnpR* gene, which is a signature that has been reported previously twice within a uropathogenic *E. coli* strain from Australia, and in association with a different class 1 integron structure [49]. A Tn*3* transposon has inserted within the *mer* module of Tn*21* with partial deletion of *merA* and *merT* and complete deletion of *merC* and *merP*. The transposon is abutted downstream of *merR* by an inward facing IS*1* insertion element. One strain, 2009-64, housed this exact structure apart from the Tn*21 tnpM*, which appears to have been lost due to an IS*26*-mediated deletion event.

The complex resistance locus (CRL) shown in Figure 3B was identified in isolates 2009-6 and 2011-69, and shares homology with the structure in Figure 3A. It bore identical IS*26* and Tn*3* insertion points, with the only major difference being a crossover event where the standard Tn*3 tnpA* gene and terminal inverted repeat have been replaced by that of Tn*1000*, a transposon originally identified in a cosmid clone of a human DNA sequence in 1995[50]. This signature was

recently identified in the sequence of an unannotated plasmid of a *Salmonella enterica* serovar Typhi strain sequenced as part of a larger study of Typhi from typhoid-endemic regions of Asia and Africa (Accession LT904889.1). This is therefore the first report of this hybrid transposon and its presence in an *E. coli* isolated in Australia. Due to the nature of Illumina sequence technology we have no confirmed sequence information downstream of Tn*3*/Tn*1000*.

Structure 3 (Fig. 3C) shares homology to the previously discussed structures. However, this CRL, present in strains 2009-38 and 2009-45, has a $bla_{\text{OXA-1}}$ gene cassette within the integron cassette array in addition to *aadA1*. Here, the Tn*21* transposon housing the class 1 integron is complete, with both the initial and terminal inverted repeats intact, and has an inward facing IS*1* flanking its *mer* end. There are two variants of this CRL in our collection, one of which contains a complete *catA1* gene downstream of Tn*21* followed by a second IS*1* element in the same orientation as the first, with the intergenic ORF identified as an acetyltransferase. This appears to be an established insertion event, with numerous reports in GenBank. In the second variant, the terminal IS*1* is inverted with consequent deletion of 476 bp of the *catA1* gene, forming a signature unique to isolates 2009-38 and 2009-45. This integron has been reported in its entirety in *Shigella dysenteriae* 1 plasmid p3099-85 (KT754164.1), *Salmonella enterica* serovar Typhimurium plasmid pUO-StVR2 (AM991977.1) and *Salmonella enterica* serovar Typhimurium strain T000240 (AP011957.1) Less than 10 SNPs were identified in comparative BLASTn alignments spanning the integron.

Sixty genomes from the SRA cohort returned adequate alignments to integron marker genes, with 25 of these appearing to possibly have only the base class 1 integron with an *aadA1* cassette but no indication of a bordering Tn*21* transposon. Eight contained an *intI1* gene but no *aadA1*, suggesting the likely presence of a class 1 integron with a different cassette array. Sixteen contained *aadA1* but no class 1 integrase; this could indicate a deletion event or more likely the presence of *aadA1* in a class 2 integron, though the *aadA1* gene can also exist independent of integron context (KX462014). Five genomes contained an unidentifiable integron structure, possibly variants of those described in the Sydney collection though it is impossible to say this definitively from read alignments against the abridged gene database used here.

Only three genomes; HVH_93_4-5851025, MOD1-EC6690 and MOD1-EC6783, contained all marker genes necessary to potentially contain integron C (Figure 3). However, within the SRA cohort, the presence of integrons A and B could not be confirmed.

All 16 strains of ST73 that were sourced from Sydney were shown to carry one or more plasmids (up to five) that ranged in size from 15 to 180 kb. Only one plasmid in each strain hybridized with the *intI1* probe (data not shown). The sizes varied greatly between 80kb and >200 kb.

Discussion

This study forms a part of wider global efforts to further understand the structure of disease-causing ST73 clones. Whole genome sequencing and maximum-likelihood phylogenetic analyses of these clones is providing important information on the community structure of ExPEC. Here we examined 16 ST73 isolates sourced from a single hospital and used sequence data sourced from the Sequence Read Archive to place these isolates into a broader global context and aid in identifying clonal lineages. Phylogenetic trees from this combined dataset, when overlaid with geographical and temporal data sourced from EnteroBase (data not shown), indicate that ST73 is globally disseminated in a manner similar to ST131[51, 52] which is currently the most studied pandemic ExPEC lineage due to the frequency of CTX-M gene carriage. However, while ST131 tends to be relatively conserved in terms of core genome, ST73 appears more variable. Analysis of locally sequenced strains and comparison to globally-sourced reads from public databases can provide context which can allow the identification of outbreak clusters with more confidence than using total SNP counts alone and may help elucidate key outbreak groups and improve public health control of disease. This is valuable as the identification of clonal groups associated with outbreaks within larger bacterial populations remains a challenge.

Characterisation of molecular signatures can also assist in the identification of outbreaks as their transfer requires physical proximity of cells. CRL including integrons and transposons are common sites of genetic rearrangement and frequently carry unique molecular signatures due to insertion elements such as IS*26*[53-55]. While all class 1 integrons in the Sydney collection are not necessarily novel, there are IS-mediated signature deletions which do not appear to have been

widely reported based on current literature such as that of the *catA1* gene. This suggests that these are local integron variants, an idea consistent with the lack of these structures in the global SRA cohort. The major representative class 1 integron described here has been reported in its entirety once within an Australian *E. coli* O2:K1:H7 ST95 strain isolated from a bloodstream infection in 2010 (unpublished data; GenBank accession CP021289.1). This integron also shares an IS*26*-mediated deletion of the Tn*21* resolvase gene *tnpR* with plasmid pUO-SeVR1 from a Spanish *Salmonella enterica* serovar Enteritidis strain sourced from a child with gastroenteritis[49, 56]. This is significant as this precise signature is likely the product of a single event. As such, a lateral transfer event is a likely explanation for the occurrence of this signature in disparate and geographically separate strains followed by changes in class 1 integron cassette content. Based on the plasmid typing and PFGE data it is likely that transfer of these integrons is being facilitated by IncF plasmids similar to pUO-SeVR1, as this is the major plasmid incompatibility type within our ST73 collection and our S1-PFGE data confirm that the class 1 integrons described here are plasmid-borne. Plasmids appear to increasingly play an important role in the mobilisation of drug resistance genes in ExPEC ST73, and their characterization relies heavily on the use of whole-genome sequencing (ideally long-read) and read-mapping technologies such as those described here.

Whole genome sequencing allows for the analysis of gene presence/absence in clinical isolates which will provide data on the importance of virulence genes in pathogenesis. The virulence profiles of strains sequenced in this study are consistent with other examinations of virulence in ST73 and in ExPEC more broadly. Genes encoding P fimbrial adhesins, the aerobactin siderophore *(iuc/iut*), and toxins hemolysin A and cytotoxic necrotizing factor 1 are not universally identified in worldwide ExPEC populations sourced from humans and animals[1, 57-59]. In previous work on ST73 isolates sourced from the UK, the prevalence of these genes/gene families was also non-universal; however, *hlyA* and *cnf1* showed substantially higher prevalence in ST73 compared with ExPEC-associated ST10, ST69 and ST95[1]. In isolates sourced from Sydney, a relatively clear association could be identified between phylogenetic groups and virulence profiles. Further *in silico* categorization of virulence profiles using global ST73 reads would provide insight into virulence patterns/groups within ST73 and ExPEC, which could potentially lead to improved response, prevention and treatment of ExPEC linked-disease.

In endemic pathogens like *E. coli*, genetic comparisons of clonal group and mobile genetic element diversity can be difficult to perform with localized populations as high numbers of closely related isolates are required for robust SNP-phylogenetic analysis and this may require the long-term collection of bacterial isolates to isolate a sufficient number of representatives. Here we used Illumina sequencing combined with SNP-phylogenetic methods to identify at least eight distinct clonal lineages in a pilot sample of 16 ST73 isolates collected from a single hospital, indicating the wealth of diversity within ST73 population sourced from highly localised sampling over an extended period (Fig. 4). Contrastingly, the diversity of mobile elements within this cohort is much less profound. Only three resistance containing class 1 integron structures were identified, all were linked to plasmids, and all showed high structural similarity. Our study is an example of how genome sequencing can provide a depth of information not available with previous molecular epidemiology methodologies that are useful in the determination of outbreak groups among ST73.

Author statements

DB performed the phylogenetic analysis, contributed to gene presence analyses, prepared most of the Figures, Tables and supplementary data and drafted multiple iterations of the manuscript. JM isolated genomic DNA, contributed to gene presence analyses, characterised structures of class 1 integrons of Sydney strains as well as generated the figure and contributed to writing the manuscript. ML constructed libraries and sequenced the Sydney ST73 strains. CV and JI performed the S1-PGFE and southern blot analyses and edited final drafts of the manuscript. AD assisted with data analysis. PRC initiated the collaboration with Sydney Adventist Hospital, screened isolates to create the strain collection included in this study and helped JM with analysis of the data presented. SD initiated and coordinated the project and drafted iterations of the final manuscript.

Funding information

This work was supported by the Australian Research Council, Linkage Grant LP150100912. This project was partly funded by the Australian Centre for Genomic Epidemiological Microbiology (Ausgem), a collaborative partnership between the NSW Department of Primary Industries and

the University of Technology Sydney. JM is a recipient of Australian Government Research Training Program Scholarships.

Acknowledgements

We acknowledge the efforts of staff from the Sydney Adventist Hospital for providing the Sydney ST73 strains and associated metadata for this study.

Ethical statement

Ethics clearance was not required.

Conflicts of interest

There are no conflicts of interest to declare.

Abbreviations

ST	Sequence type
ExPEC	Extraintestinal *Escherichia coli*
SNP	Single nucleotide polymorphism
S1-PFGE	S1 Nuclease - Pulsed Field Gel Electrophoresis
CS	Conserved segments
UPEC	Uropathogenic *E. coli*
NMEC	Neonatal meningitis-causing *E. coli*
APEC	Avian pathogenic *E. coli*
UTI	Urinary tract infection

VAG	Virulence-associated gene
MDR	Multiple drug resistant
SAN	Sydney Adventist Hospital
LB	Luria-Bertani
AMP	Ampicillin
AMC	Amoxicillin-clavulanic acid
LEX	Cephalexin
CIP	Ciprofloxacin
GEN	Gentamicin
NIT	Nitrofurantoin
NOR	Norfloxacin
TMP	Trimethoprim

Table 1: Phenotypic antimicrobial resistance

Patient	AMP	AMC	LEX	CIP	GEN	NIT	NOR	TMP
2009_3	R	R	S	S	S	S	S	S
2009_5	R	R	S	S	S	S	S	S
2009_6	R	R	S	S	S	S	-	S
2009_8	R	S	S	S	S	S	-	S
2009_13	R	S	S	S	S	S	S	S
2009_14/16	R	S	S	S	S	S	-	S
2009_21	R	S	S	S	S	S	-	S
2009_38	R	S	S	S	S	S	S	S
2009_44	R	-	S	S	S	S	S	S
2009_45	R	S	S	S	S	S	S	S
2010_64	R	S	S	S	S	S	-	S
2011_69	R	S	S	S	S	R	S	S
2011_82	S	-	S	S	S	S	S	R
2011_93/98	R	S	S	S	S	S	-	S

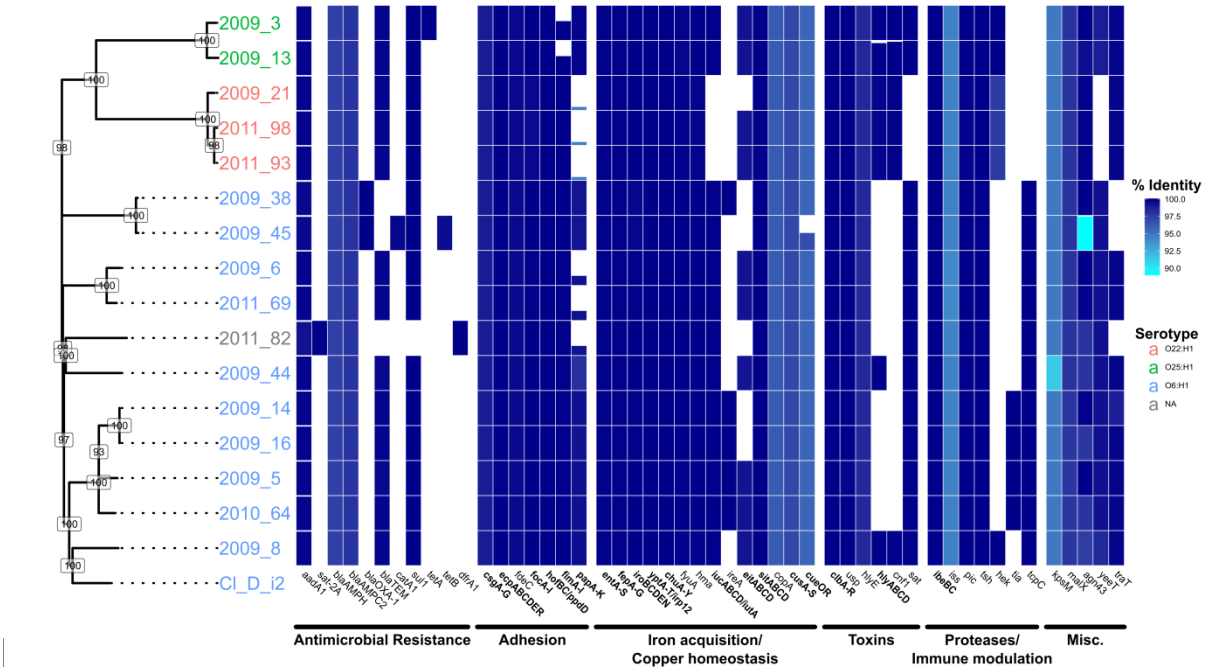


Figure 1: A SNP-dervied phylogenetic tree of the Sydney ST73 strains sequenced in this study compared with antimicrobial resistance and virulence profiles. Bootstrap values based on 100 replicate trees are shown at labelled nodes. Isolate serotypes as determined by *in silico* serotyping are shown by coloured tip labels. Antibiotic resistance and virulence gene/gene-family presence (blue) or absence (white) is shown by the linked bar-graph/heatmap. Percentage identity of BLAST matches is indicated by heatmap shade with darker shades representing higher identity. BLAST match coverage is represented by tile height with solid tiles representing 100% coverage. For gene families (x-axis; bold) tile height represents total BLAST match coverage of all gene family members and shows completeness of the gene family. Where single genes are indicated (x-axis; plain text), bar height represents BLAST match coverage of the gene.

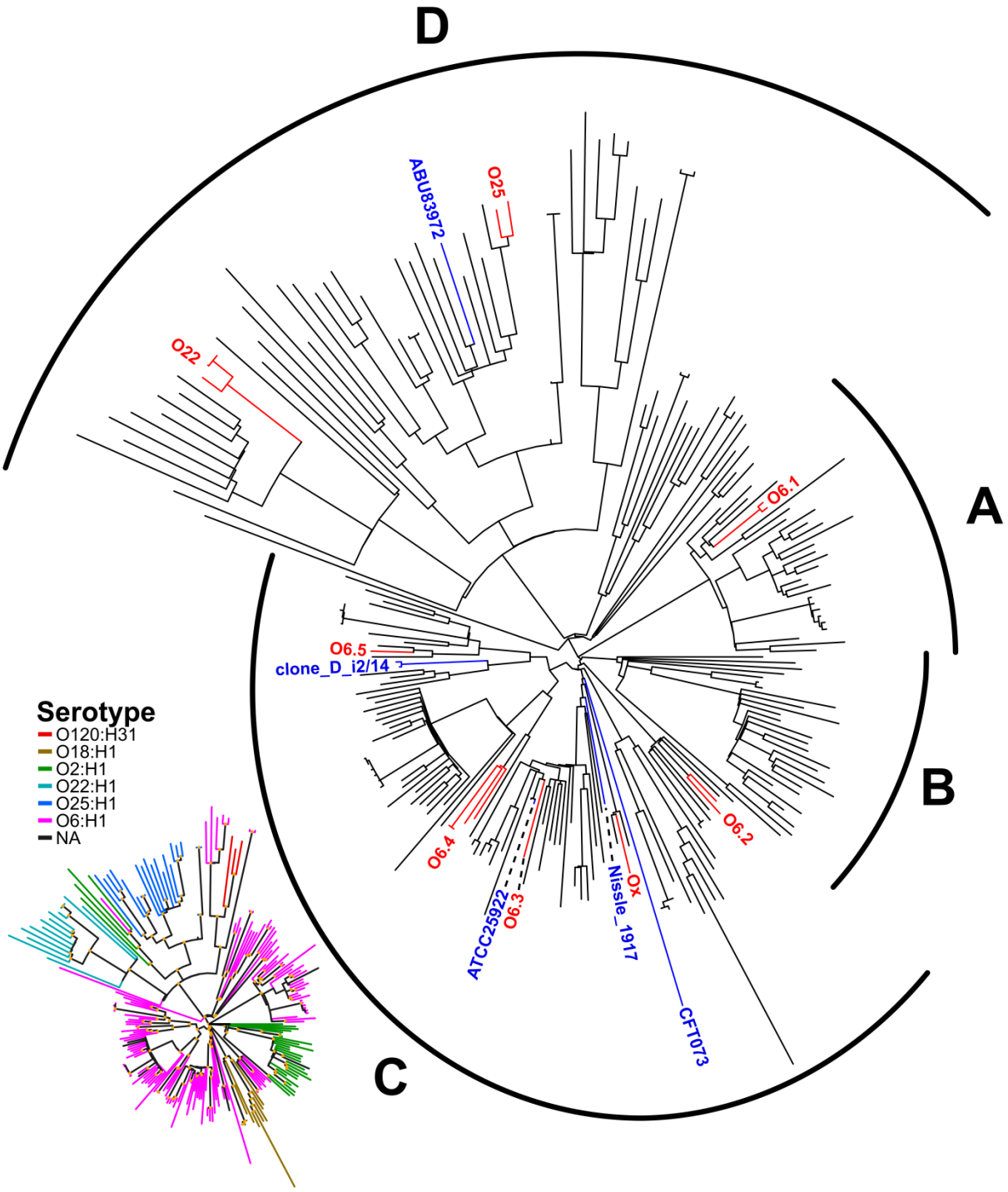


Figure 2: SNP-based maximum-likelihood phylogram of 226 ST73 strains. A more detailed tree with branch support values and tip labels can be found in Figure S2. ST73 isolates separate into 4 distinct groups, labelled A-D, which correlate well with *in silico* serotyping (insert). ST73 isolates sequenced in this study cluster into 8 distinct groups, shown in red, high-quality complete ST73 genomes are shown in blue. Trees were constructed using 18,426 SNPs identified by read mapping to the clone D i2 reference sequence, reduced from 27,568 SNPs by filtering of recombination regions.

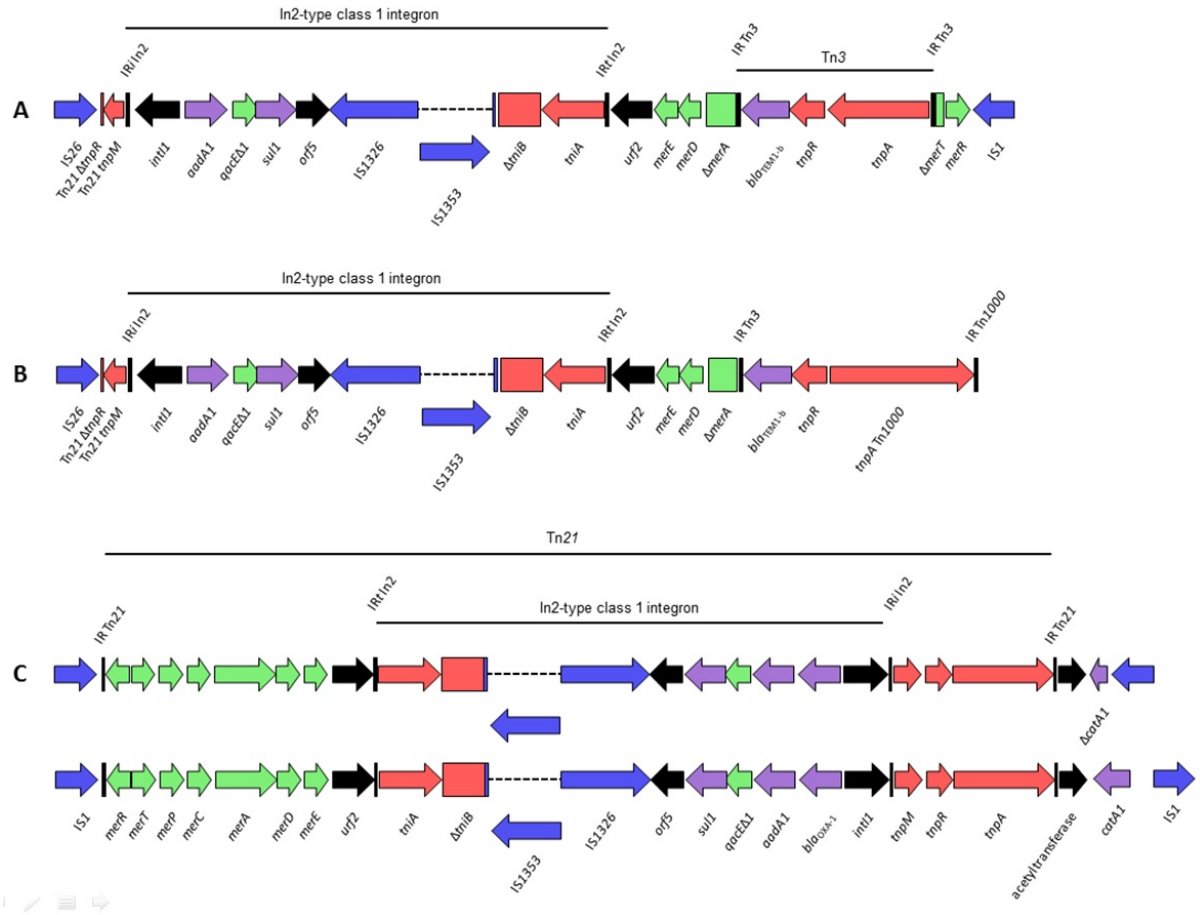


Figure 3: Schematic representations of integron structures found within this collection.

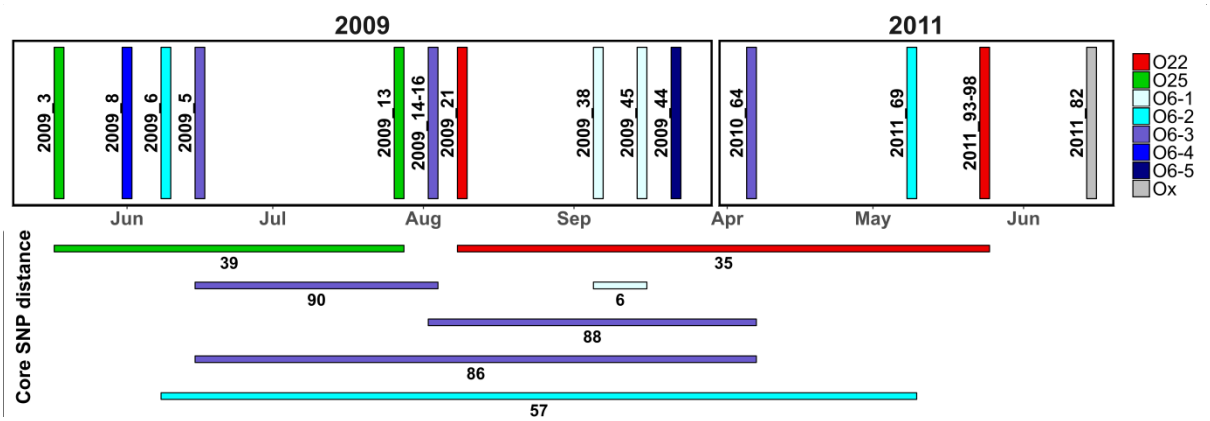


Figure 4. Epidemic curve for ST73 isolates sequenced in this study. Isolates are coloured to match groups identified by comparison to isolates from the Sequence Read Archive. Core SNP distances between samples of the same group are shown with coloured horizontal lines.

References

1. Gibreel TM, Dodgson AR, Cheesbrough J et al. Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *J Antimicrob Chemother* 2012; **67**: 346-56.
2. Kim S, Sung JY, Cho HH et al. Characteristics of the Molecular Epidemiology of CTX-M-Producing *Escherichia coli* Isolated from a Tertiary Hospital in Daejeon, Korea. *J Microbiol Biotechnol* 2016; **26**: 1643-9.
3. Manges AR, Johnson JR. Reservoirs of Extraintestinal Pathogenic *Escherichia coli*. *Microbiol Spectr* 2015; **3**.
4. Olesen B, Scheutz F, Menard M et al. Three-Decade Epidemiological Analysis of *Escherichia coli* O15:K52:H1. *J Clin Microbiol* 2009; **47**: 1857-62.
5. Yun KW, Kim DS, Kim W et al. Molecular typing of uropathogenic *Escherichia coli* isolated from Korean children with urinary tract infection. *Korean Journal of Pediatrics* 2015; **58**: 20-7.
6. Johnson JR, Thuras P, Johnston BD et al. The Pandemic H30 Subclone of *Escherichia coli* Sequence Type 131 Is Associated With Persistent Infections and Adverse Outcomes Independent From Its Multidrug Resistance and Associations With Compromised Hosts. *Clin Infect Dis* 2016; **62**: 1529-36.
7. Pitout JD. Extraintestinal Pathogenic Escherichia coli: A Combination of Virulence with Antibiotic Resistance. *Front Microbiol* 2012; **3**: 9.
8. Wirth T, Falush D, Lan R et al. Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol Microbiol* 2006; **60**: 1136-51.
9. Phan MD, Peters KM, Sarkar S et al. The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. *PLoS Genet* 2013; **9**: e1003834.
10. de Kraker ME, Jarlier V, Monen JC et al. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. *Clin Microbiol Infect* 2013; **19**: 860-8.
11. Alhashash F, Wang X, Paszkiewicz K et al. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73: high levels of genomic and plasmid diversity in causative isolates. *J Antimicrob Chemother* 2016; **71**: 339-43.

12. de Souza da-Silva AP, de Sousa VS, Martins N et al. *Escherichia coli* sequence type 73 as a cause of community acquired urinary tract infection in men and women in Rio de Janeiro, Brazil. *Diagn Microbiol Infect Dis*; **88**: 69-74.
13. Miajlovic H, Mac Aogáin M, Collins CJ et al. Characterization of *Escherichia coli* bloodstream isolates associated with mortality. *J Med Microbiol* 2016; **65**: 71-9.
14. Yahiaoui M, Robin F, Bakour R et al. Antibiotic Resistance, Virulence, and Genetic Background of Community-Acquired Uropathogenic Escherichia coli from Algeria. *Microb Drug Resist* 2015; **21**: 516-26.
15. Clermont O, Gordon D, Denamur E. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* 2015; **161**: 980-8.
16. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin Microbiol Infect* 2014; **20**: 380-90.
17. Bengtsson S, Naseer U, Sundsfjord A et al. Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *J Antimicrob Chemother* 2012; **67**: 69-73.
18. Kahlmeter G, Åhman J, Matuschek E. Antimicrobial Resistance of *Escherichia coli* Causing Uncomplicated Urinary Tract Infections: A European Update for 2014 and Comparison with 2000 and 2008. *Infectious Diseases and Therapy* 2015; **4**: 417-23.
19. Gillings MR, Gaze WH, Pruden A et al. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* 2015; **9**: 1269-79.
20. Bell SM, Pham JN, Rafferty DL et al. *Antibiotic susceptibility tesing by the CDS method: A manual for medical and veterinary laboratories 2016*. Kogarah, Australia: South Eastern Area Laboratory Services, 2016.
21. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015; **31**: 587-9.
22. Bankevich A, Nurk S, Antipov D et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**: 455-77.
23. Inouye M, Dashnow H, Raven LA et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014; **6**: 90.

24. Chen L, Yang J, Yu J et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005; **33**: D325-8.

25. Gupta SK, Padmanabhan BR, Diene SM et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014; **58**: 212-20.

26. Minogue TD, Daligault HA, Davenport KW et al. Complete Genome Assembly of *Escherichia coli* ATCC 25922, a Serotype O6 Reference Strain. *Genome Announc* 2014; **2**: e00969-14.

27. Reeves PR, Liu B, Zhou Z et al. Rates of Mutation and Host Transmission for an *Escherichia coli* Clone over 3 Years. *PLoS One* 2011; **6**: e26907.

28. Reister M, Hoffmeier K, Krezdorn N et al. Complete genome sequence of the gram-negative probiotic *Escherichia coli* strain Nissle 1917. *J Biotechnol* 2014; **187**: 106-7.

29. Zdziarski J, Brzuszkiewicz E, Wullt B et al. Host imprints on bacterial genomes--rapid, divergent evolution in individual patients. *PLoS Pathog* 2010; **6**: e1001078.

30. Welch RA, Burland V, Plunkett G, 3rd et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 2002; **99**: 17020-4.

31. Dallman TJ, Byrne L, Ashton PM et al. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin Infect Dis* 2015; **61**: 305-12.

32. Dunitz MI, Coil DA, Jospin G et al. Draft Genome Sequences of *Escherichia coli* Strains Isolated from Septic Patients. *Genome Announc* 2014; **2**.

33. Earle SG, Wu CH, Charlesworth J et al. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol* 2016; **1**: 16041.

34. Kluytmans-van den Bergh MF, Rossen JW, Bruijning-Verhagen PC et al. Whole-Genome Multilocus Sequence Typing of Extended-Spectrum-Beta-Lactamase-Producing Enterobacteriaceae. *J Clin Microbiol* 2016; **54**: 2919-27.

35. Pesesky MW, Hussain T, Wallace M et al. KPC and NDM-1 genes in related Enterobacteriaceae strains and plasmids from Pakistan and the United States. *Emerg Infect Dis* 2015; **21**: 1034-7.

36. Roach DJ, Burton JN, Lee C et al. A Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota. *PLoS Genet* 2015; **11**: e1005413.

37. Salipante SJ, Roach DJ, Kitzman JO et al. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res* 2015; **25**: 119-28.

38. Toro M, Cao G, Rump L et al. Genome Sequences of 64 Non-O157:H7 Shiga Toxin-Producing *Escherichia coli* Strains. *Genome Announc* 2015; **3**.

39. Barton BM, Harding GP, Zuccarelli AJ. A general method for detecting and sizing large plasmids. *Anal Biochem* 1995; **226**: 235-40.

40. Partridge SR, Zong Z, Iredell JR. Recombination in IS26 and Tn2 in the evolution of multiresistance regions carrying blaCTX-M-15 on conjugative IncF plasmids from *Escherichia coli*. *Antimicrob Agents Chemother* 2011; **55**: 4971-8.

41. Koeleman JG, Stoof J, Van Der Bijl MW et al. Identification of epidemic strains of *Acinetobacter baumannii* by integrase gene PCR. *J Clin Microbiol* 2001; **39**: 8-13.

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

43. Croucher NJ, Harris SR, Fraser C et al. Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011; **331**: 430-4.

44. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312-3.

45. Nesta B, Spraggon G, Alteri C et al. FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *MBio* 2012; **3**.

46. Ladomersky E, Petris MJ. Copper tolerance and virulence in bacteria. *Metallomics* 2015; **7**: 957-64.

47. Baker-Austin C, Wright MS, Stepanauskas R et al. Co-selection of antibiotic and metal resistance. *Trends Microbiol* 2006; **14**: 176-82.

48. Nakaya R, Nakamura A, Murata Y. Resistance transfer agents in Shigella. *Biochem Biophys Res Commun* 1960; **3**: 654-9.

49. Garcia V, Garcia P, Rodriguez I et al. The role of IS26 in evolution of a derivative of the virulence plasmid of *Salmonella enterica* serovar Enteritidis which confers multiple drug resistance. *Infect Genet Evol* 2016; **45**: 246-9.

50. Broom JE, Hill DF, Hughes G et al. Sequence of a transposon identified as Tn1000 (gamma delta). *DNA Seq* 1995; **5**: 185-9.
51. Petty NK, Ben Zakour NL, Stanton-Cook M et al. Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A* 2014; **111**: 5694-9.
52. Price LB, Johnson JR, Aziz M et al. The Epidemic of Extended-Spectrum-β-Lactamase-Producing *Escherichia coli* ST131 Is Driven by a Single Highly Pathogenic Subclone, H30-Rx. *mBio* 2013; **4**: e00377-13.
53. Dawes FE, Kuzevski A, Bettelheim KA 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**: e12754.
54. 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**: e0115781.
55. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. *Int J Antimicrob Agents* 2018; **52**: 430-5.
56. Rodriguez I, Rodicio MR, Herrera-Leon S et al. Class 1 integrons in multidrug-resistant non-typhoidal *Salmonella enterica* isolated in Spain between 2002 and 2004. *Int J Antimicrob Agents* 2008; **32**: 158-64.
57. Cunha MPV, Saidenberg AB, Moreno AM et al. Pandemic extra-intestinal pathogenic *Escherichia coli* (ExPEC) clonal group O6-B2-ST73 as a cause of avian colibacillosis in Brazil. *PLoS One* 2017; **12**: e0178970.
58. Liu X, Thungrat K, Boothe DM. Multilocus Sequence Typing and Virulence Profiles in Uropathogenic *Escherichia coli* Isolated from Cats in the United States. *PLoS One* 2015; **10**: e0143335.
59. Yun KW, Kim HY, Park HK et al. Virulence factors of uropathogenic *Escherichia coli* of urinary tract infections and asymptomatic bacteriuria in children. *J Microbiol Immunol Infect* 2014; **47**: 455-61.

Chapter 5: Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis

Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis

Jessica McKinnon [a,1], Piklu Roy Chowdhury [a,b,1], Steven P. Djordjevic [a]

[a] *ithree Institute, University of Technology Sydney, P.O. Box 123, Broadway, Sydney, NSW 2007, Australia*

[b] *Department of Primary Industries, Elizabeth Macarthur Agriculture Institute, PMB 4008, Camden, NSW 2567, Australia*

[1] These two authors contributed equally to this article.

Abstract

Sequence type 58 (ST58) phylogroup B1 *Escherichia coli* have been isolated from a wide variety of mammalian and avian hosts but are not noted for their ability to cause serious disease in humans or animals. Here we determined the genome sequences of two multidrug-resistant *E. coli* ST58 strains from urine and blood of one patient using a combination of Illumina and Single Molecule, Real-Time (SMRT) sequencing. Both ST58 strains were clonal and were characterised as serotype O8:H25, phylogroup B1 and carried a complex resistance locus/loci (CRL) that featured an atypical class 1 integron with a *dfrA5* (trimethoprim resistance) gene cassette followed by only 24 bp of the 3′-CS. CRL that carry this particular integron have been described previously in *E. coli* from cattle, pigs and humans in Australia. The integron abuts a copy of Tn*6029*, an IS*26*-flanked composite transposon encoding bla_{TEM}, *sul2* and *strAB* genes that confer resistance to ampicillin, sulfathiazole and streptomycin, respectively. The CRL resides within a novel Tn*2610*-like hybrid Tn*1721*/Tn*21* transposon on an IncF, ColV plasmid (pSDJ2009-52F) of 138 553 bp that encodes virulence associated genes implicated in life-threatening extraintestinal pathogenic *E. coli* (ExPEC) infections. Notably, pSDJ2009-52F shares high sequence identity with pSF-088-1, a plasmid reported in an *E. coli* ST95 strain from a patient with blood sepsis from a hospital in San Francisco. These data suggest that extraintestinal infections caused by *E. coli* carrying ColV-like plasmids, irrespective of their phylogroup or ST, may pose a potential threat to human health, particularly to the elderly and immunocompromised.

Keywords: *Escherichia coli* ST58, ExPEC, Virulence plasmid, Complex resistance region, Tn*6029*, Tn*2610*

1. Introduction

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are a phylogenetically diverse group of *E. coli* that have acquired the ability to colonise anatomical sites distinct from the gastrointestinal tract, including the urinary tract, brain and spinal cord, soft tissue sites and bone. ExPEC are classified into three subtypes, including uropathogenic *E. coli* (UPEC), neonatal meningitis-associated *E. coli* and avian pathogenic *E. coli*[1]. ExPEC also cause nosocomial bloodstream infections in hospitals and nursing homes and are often responsible for respiratory infections and bacteraemia in long-term hospitalised patients. ExPEC have been identified in humans, food animals, retail meat products, companion animals and effluent and there is a significant body of circumstantial evidence to suggest that a subset of ExPEC are zoonotic pathogens[2-4].

Our understanding of the evolution and spread of multiple antimicrobial resistance in ExPEC has been influenced by efforts to gauge the prevalence of *E. coli* expressing extended-spectrum β-lactamases, Amp C β-lactamases, and plasmid-mediated quinolone and fluoroquinolone resistance. Knowledge of the extent of carriage of antimicrobial resistance genes is likely to be underestimated because many studies restrict phenotypic screening to antimicrobials that are of clinical relevance for infection control. As a consequence, resistance to older first-generation antimicrobials, many of which are still used in veterinary medicine and food production more broadly, particularly ampicillin, sulfonamides, tetracycline and trimethoprim, is often ignored.

Insertion sequence IS*26* plays a critical role in the capture and dissemination of broad classes of antimicrobial resistance genes, including genes encoding resistance to all last-line antimicrobial agents. IS*26* is widespread in commensal and pathogenic *E. coli* from food animals in Australia[5, 6] and it often localises to sites that are within close genetic proximity to class 1 integrons, facilitating the formation of complex resistance loci (CRL) on transposons and plasmids[5, 7-13]. IS*26* can promote the generation of hybrid plasmids carrying antimicrobial resistance and virulence genes[14] as well as influence plasmid fitness by deleting regions of DNA that encode genes whose expression incurs a cost to the host[15]. Furthermore, CRL containing existing copies of IS*26* are targeted by mobile genetic cargo flanked by IS*26*, enabling host bacteria to adapt rapidly to antimicrobial selection pressure. Clearly it is important to understand how CRL encoding resistance to first-generation antimicrobials, and the mobile elements that carry them,

have evolved and to determine how widespread they are in commensal and pathogenic *E. coli* as well as other Gram-negative pathogens that are exposed to antimicrobial selection pressures in human healthcare, food animal production settings and the environment.

Here we characterised two UPEC isolates from an elderly patient collected 4 days apart. Strain 2009-49 was from midstream urine and strain 2009-52 was from blood. The aim of this study was to use whole-genome sequencing (WGS) to determine the genetic relationship they shared with one another and with isolates belonging to the same multilocus sequence typing (MLST) sequence type, their repertoire of antimicrobial resistance and virulence-associated genes, and the plasmids they carry.

Materials and Methods

2.1. Bacterial strains

Both ST58 *E. coli* strains from this study were sourced from Sydney Adventist Hospital (Sydney, NSW, Australia) in 2009. Strain 2009-49 was cultured from the midstream urine of an elderly patient presenting with a urinary tract infection (UTI), and strain 2009-52 was cultured from a blood sample obtained from the same patient 4 days later. The pathology report indicated that the patient had elevated leukocyte and red blood cell counts and a high bacterial colony count. The urinary tract strain 2009-49 displayed resistance to ampicillin, trimethoprim and nitrofurantoin, whilst the blood strain was recorded as being resistant to ampicillin.

2.2. Calibrated dichotomous sensitivity (CDS) testing

Antimicrobial susceptibility testing was performed using the established CDS protocol. Strains 2009-49 and 2009-52 as well as three reference strains (*E. coli* ACM 5185 and 5186 and *Pseudomonas aeruginosa* ACM 5189 137) were grown overnight on lysogeny broth agar (Sigma-Aldrich; Sydney, Australia) plates at 37 °C. Antimicrobial disks were applied to Sensitest® (Oxoid product from Thermo Fisher Scientific Australia; Scoresby, Victoria, Australia) plates at a maximum of six disks per plate using a 6-Cartridge Disc Dispenser (Oxoid). The strains were tested against 20 antimicrobials, including ampicillin (25 μg),

apramycin (15 μg), Augmentin® (amoxicillin/clavulanic acid; 6 μg), azithromycin (15 μg), cefotaxime (5 μg), cefoxitin (30 μg), cefalexin (100 μg), chloramphenicol (30 μg), ciprofloxacin (2.5 μg), gentamicin (10 μg), imipenem (10 μg), kanamycin (50 μg), nalidixic acid (30 μg), neomycin (30 μg), nitrofurantoin (200 μg), streptomycin (25 μg), sulfafurazole (300 μg), tetracycline (10 μg), Timentin® (ticarcillin/clavulanic acid; 85 μg) and trimethoprim (5 μg).

2.3. Whole-genome sequencing and analyses

WGS was performed using an Illumina® MiSeq platform (Illumina Inc.; San Diego, California, United States) at the ithree Institute, University of Technology Sydney (Sydney, NSW, Australia) following published protocols, and raw reads were assembled *de novo* using the A5-miseq pipeline[16]. MLST was performed electronically using the *E. coli* Achtman MLST scheme available at https://pubmlst.org/. Plasmid classification was determined by submission of whole genomes to PlasmidFinder 1.2 (https://cge.cbs.dtu.dk/services/PlasmidFinder/).

Single Molecule, Real-Time (SMRT) sequencing was performed on strain 2009-52 using a PacBio RSII instrument (Pacific Biosciences; Menlo Park, California, United States) and HGAP assembly at the Ramaciotti Centre for Genomics at the University of New South Wales (Sydney, Australia). The sequence was polished using Pilon (https://github.com/broadinstitute/pilon). A phylogenetic tree was constructed with 326 complete genome sequences of *E. coli* obtained from Enterobase (https://enterobase.warwick.ac.uk/species/index/ecoli). PhyloSift v.1.0 (http://phylosift.wordpress.com/) was used to undertake the phylogenetic analysis and the output was manipulated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). PhyloSift analysis was used to determine the genome sequence most closely related to ST58 strains 2009-49 and 2009-52.

2.4. Single nucleotide polymorphisms (SNPs)

Parsnp from the Harvest Tools suite (https://harvest.readthedocs.io/en/latest/content/harvest-tools.html) was used to identify SNPs in the core genomes of strains 2009-49 (MiSeq) and 2009-52 (PacBio) by aligning the genomes with the most closely related complete reference genome (SCK30-22) available in Enterobase. x flag was implemented to filter out recombination events,

which resulted in alignment of 82.2% of the three genomes. Gingr[17] was used to visualise the SNP locations in the .ggr file and the matching co-ordinates confirmed from the exported .vcf file. The genes affected by SNPs identified in the process were thereafter confirmed by BLASTn analysis and were subjected to SIFT (http://sift.jcvi.org/www/SIFT_seq_submit2.html) analysis to assess functional alterations in protein function.

2.5. Nucleotide sequence accession nos.

Strains 2009-49 and 2009-52 and plasmid pSDJ2009-52F have been submitted to GenBank under accession nos. **NXEP00000000**, **NXEO00000000** and **MH195200**, respectively.

3. Results and Discussion

UPEC strain 2009-49 (urine isolate) and 2009-52 (blood isolate) were sourced from the same patient within a narrow time frame (4 days). Available metadata indicated that both strains were resistant to ampicillin, trimethoprim and nitrofurantoin. CDS assays revealed that both strains showed reduced susceptibility to ampicillin, trimethoprim, sulfafurazole, tetracycline and streptomycin. In silico MLST analysis identified both strains as ST58, a sequence type that resides within the ST155 clonal complex. Both strains were Clermont phylotype B1, were serotyped in silico as O8:H25, and carried plasmid incompatibility markers F and FIB. Strain 2009-52 also carried an I2 replicon.

3.1. Whole-genome sequencing and comparative genomic analyses

Assembly statistics for the two genomes are presented in Supplementary Table S1. A proportional phylogenetic cladogram generated using PhyloSift, comprising genome sequences of strains 2009-49 and 2009-52 as well as 84 *E. coli* ST58 genomes of human and animal origin from Enterobase, is shown in Fig. 1. A more comprehensive phylogenetic tree of ST58 genomes is shown in Supplementary Fig. S1. Strains 2009-49 and 2009-52 clustered together with a node confidence value of 1. The closest relative to the Sydney ST58 strains was strain SCK30-22, a

ST58, serotype O8:H25 *E. coli* sourced from an undisclosed human infection from the Netherlands. Strain SK30-22 was used to extract the core genome of the two ST58 strains in this study and to identify SNPs between them. Only four SNPs, all but one (*fimG*; S167R) of which were silent, were identified using Parsnp and Gingr from the Harvest Tools suite. Three SNPs found in strain 2009-52 in comparison with strain 2009-49 and the reference strain were located in DNA ligase B, intergenic sequence and the *fimG* gene, respectively. One SNP was identified within a predicted transposase in strain 2009-49. The low number of SNPs is consistent with the two Sydney ST58 strains being clonal. SMRT sequencing showed that strain 2009-52 had a chromosome comprising 5 012 036 nucleotides and two plasmids: IncI2 plasmid pSDJ2009-52I2 (215 832 bp); and IncF plasmid pSDJ2009-52F (138 553 bp).

A whole-genome BLASTn analysis using the genome of strain 2009-52 as the query and 2009-49 (Supplementary Fig. S2) identified a 215 832 nucleotide (nt) plasmid (pSDJ2009-52I2) scaffold (Supplementary Fig. S3) in strain 2009-52 but not in strain 2009-49. pSDJ2009-52I2 carries a putative type IV secretion system (T4SS) (*virD4*, *virB1*, *virB2*, *virB4–B6* and *vir8–11*), phage-related genes, pilin genes (*pilN*, *pilO*, *pilP*, *pilQ*, *pilR*, *pilS*, *pilT*, *pilU* and *pilV*) and 36 hypothetical genes. Coding sequences in pSDJ2009-52I2 are shown in Supplementary Table S2.

3.2. Virulence-associated genes (VAGs) in ST58 strains

Strains 2009-49 and 2009-52 carried fimbrial adhesins including *bmaE*, *fimH*, *gafD*, *csgG* and *yfcV*, multiple iron acquisition operons, several copper resistance genes, toxins including *hlyE*, *hlyF* and *hek*, and various membrane-associated pumps and proteases (Table 1).

3.3. Characterisation of the complex resistance locus in the Sydney ST58 strains

Both strain 2009-49 and 2009-52 possess an atypical class 1 integron carrying a *dfrA5* (trimethoprim resistance) gene cassette missing all but 24 bp of the 3′-CS. Genetic signatures such as these can be useful to track plasmids and transposons[5, 10]. PCR with a forward primer in *intI1* and a reverse primer in IS*26*, developed previously to characterise atypical class 1 integrons[5], generated an amplicon of 848 bp. An identical 848-bp signature is observed in

derivate Tn*21*-associated CRL of multidrug-resistant (MDR) O26 enterohaemorrhagic *E. coli* (EHEC) from a human and serologically diverse atypical enteropathogenic *E. coli* (EPEC) from Australian cattle[5, 12, 13] and in commensal *E. coli* from swine[6]. The presence of the same signature in UPEC suggests that the Tn*21* transposon, and the plasmids that carry it, are disseminated widely in animal production systems in Australia. The molecular signature is created by an IS*26*-mediated deletion caused by insertion of IS*26*[5] or IS*26*-flanked transposons such as Tn*6029*, into the backbone of mercury resistance transposons belonging to the Tn*3* family[6, 12, 13]. Tn*6029* harbours the genes $bla_{\text{TEM-1}}$ (ampicillin resistance), *strAB* (streptomycin resistance) and *sul2* (sulfathiazole resistance). In EHEC and EPEC that carry the same molecular signature, Tn*6029*, or the related transposon Tn*6026*, abuts the atypical class 1 integron precisely 24 bp into the 3′-CS. In strains 2009-49 and 2009-52 the mercury resistance module of Tn*21* flanks one arm of the resistance region CRL, whilst the other arm comprises a hybrid version of Tn*1721*/Tn*21* that carries the tetracycline resistance gene *tetA*(A), followed by the transposition module of Tn*21* (Fig. 2B). The crossover region resides 373 nt upstream of the class 1 integron IR*i*, which notably has been reported previously in a clinical UPEC strain from Sydney (GenBank accession no. **HM999791.1**). The CRL is sufficient to encode resistance to ampicillin, streptomycin, sulfonamides, tetracycline and trimethoprim, a pattern of resistance not uncommon in Australian food animal production systems[5, 18, 19]. Nitrofurantoin resistance may be attributed to mutations or deletions in the chromosomal *nfsA* and *nfsB* genes. In both strains, *nfsA* was missing the terminal 302 nt and *nfsB* contained 10 SNPs compared with that of the nitrofurantoin sensitive (wild-type) strain K-12 and are sufficient to engender an $nfsAB^{+}$ genotype in strains 2009-49 and 2009-52.

3.4. pSDJ2009-52F carries the Tn*1721*/Tn*21* hybrid transposon

The hybrid Tn*1721*/Tn*21* transposon is located in the *repFIIa* gene in pSDJ2009-52F and is flanked by 5-bp direct repeats indicating that it, or a progenitor of it, transposed to this site. pSDJ2009-52F carries ExPEC VAGs that have a role in the acquisition of iron including *iutA*, *iucDCBA*, *shiF*, *sitDCBA*, *hlyF*, *etsABC*, *iroBCDE* and *iroN*, as well as *ompT*, *iss* and *cvaC* genes (Table 1). Both Sydney ST58 strains carry identical copies of pSDJ2009-52F. BLASTn analysis shows that the sequence flanking either side of hybrid transposon shares significant

nucleotide sequence identity with pSF-088-1 (accession no. **CP012636.1**), a plasmid found in a ST95 strain of *E. coli* recovered from a patient with a bloodstream infection at San Francisco General Hospital (San Francisco, CA) (Fig. 2A). ST95 is a sequence type linked with ExPEC disease in poultry and humans. pSF-088-1 but not pSDJ2009-52F contains the full *tra* region and lacks the *tet* genes of Tn1721. Notably, pSF-088-1 carries the same atypical class 1 integron found in the uropathogenic Sydney ST58 strains.

4. Conclusion

Here we describe the isolation and sequence analysis of two *E. coli* ST58 phylotype B1 strains from the urine and blood of the same patient. Both ST58 strains carry pSDJ2009-52F, an IncF plasmid harbouring a CRL encoding multiple antimicrobial resistance genes, IS*26* and a variety of VAGs that are typical of ColV plasmids. ColV plasmids have been implicated in ExPEC infections in humans and animals, often causing urinary, neurological and systemic disease. pSDJ2009-52F encodes resistance to multiple classes of antimicrobials, the genes of which are housed within a Tn*2610*-like hybrid Tn*1721*/Tn*21* transposon.

Escherichia coli ST58 has been isolated from mammalian and avian hosts, including European rooks, humans, cats, camels and cattle. In Enterobase (https://enterobase.warwick.ac.uk/species/index/ecoli), ST58 strains are described from a range of sources, including humans, pigs, cows, horses, retail meat, animal feed, dogs, turkeys and various other birds, a prairie dog, goats, sheep, spinach, water and soil, and from various geographic locations, including Australia, Ireland, France, Denmark, Tanzania, Ethiopia, Japan, Poland, Thailand, Sweden, Belgium, Canada, Spain, the Netherlands, USA, Brazil, Germany, the UK, China and Bangladesh (accessed 6 June 2018). To our knowledge, this is the first description of an MDR *E. coli* ST58 strain associated both with UTI and blood sepsis. These studies provide an example of how acquisition of a single plasmid can impart an arsenal of virulence and antimicrobial resistance genes to a commensal *E. coli* with a wide host range and suggests that acquisition of the IncF plasmid pSD2009-52F is significant in the evolution of the Sydney ST58 strains and, potentially, their ability to cause UTI and urosepsis.

Acknowledgments

The authors thank Nola Hitchick and Natalie Miller for providing the *Escherichia coli* ST58 strains from the Sydney Adventist Hospital (Sydney, NSW, Australia). The authors also thank Ethan Wyrsch for assistance with bioinformatics analyses, and Jaqueline Melvold for assistance with analysis of hypothetical proteins.

Funding

This work was supported by Ausgem, a collaborative partnership between the New South Wales Department of Primary Industries' Elizabeth Macarthur Agricultural Institute (EMAI) and the ithree Institute at the University of Technology Sydney (UTS). JM is the recipient of an Australian Government Research Training Program Scholarship.

Competing interests

None declared.

Ethical approval

Not required.

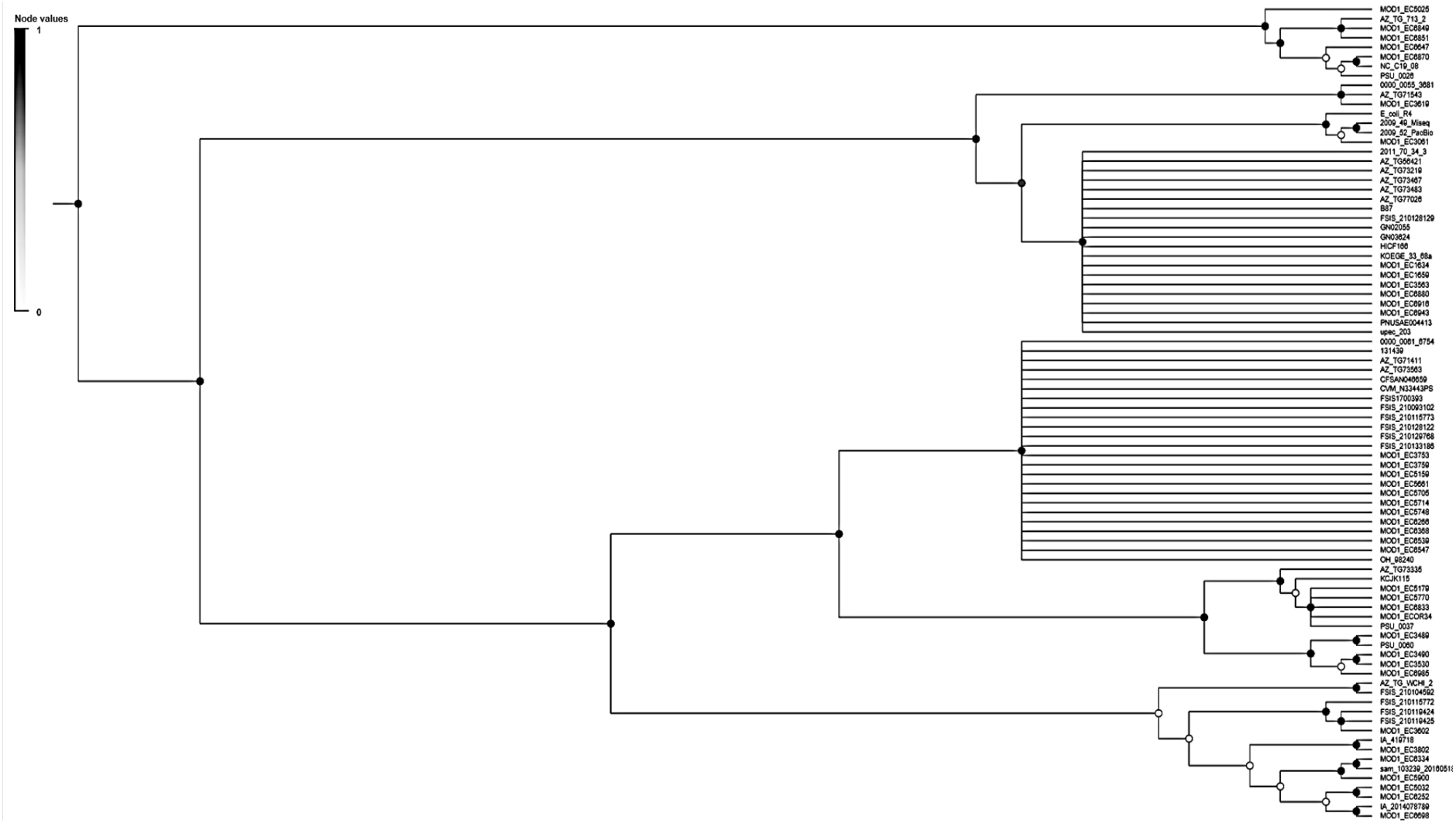


Figure 1: Proportional cladogram of all *Escherichia coli* sequence type 58 (ST58) genomes from Enterobase with complete metadata according to the selection criteria. A black circle indicates a node value of 1.

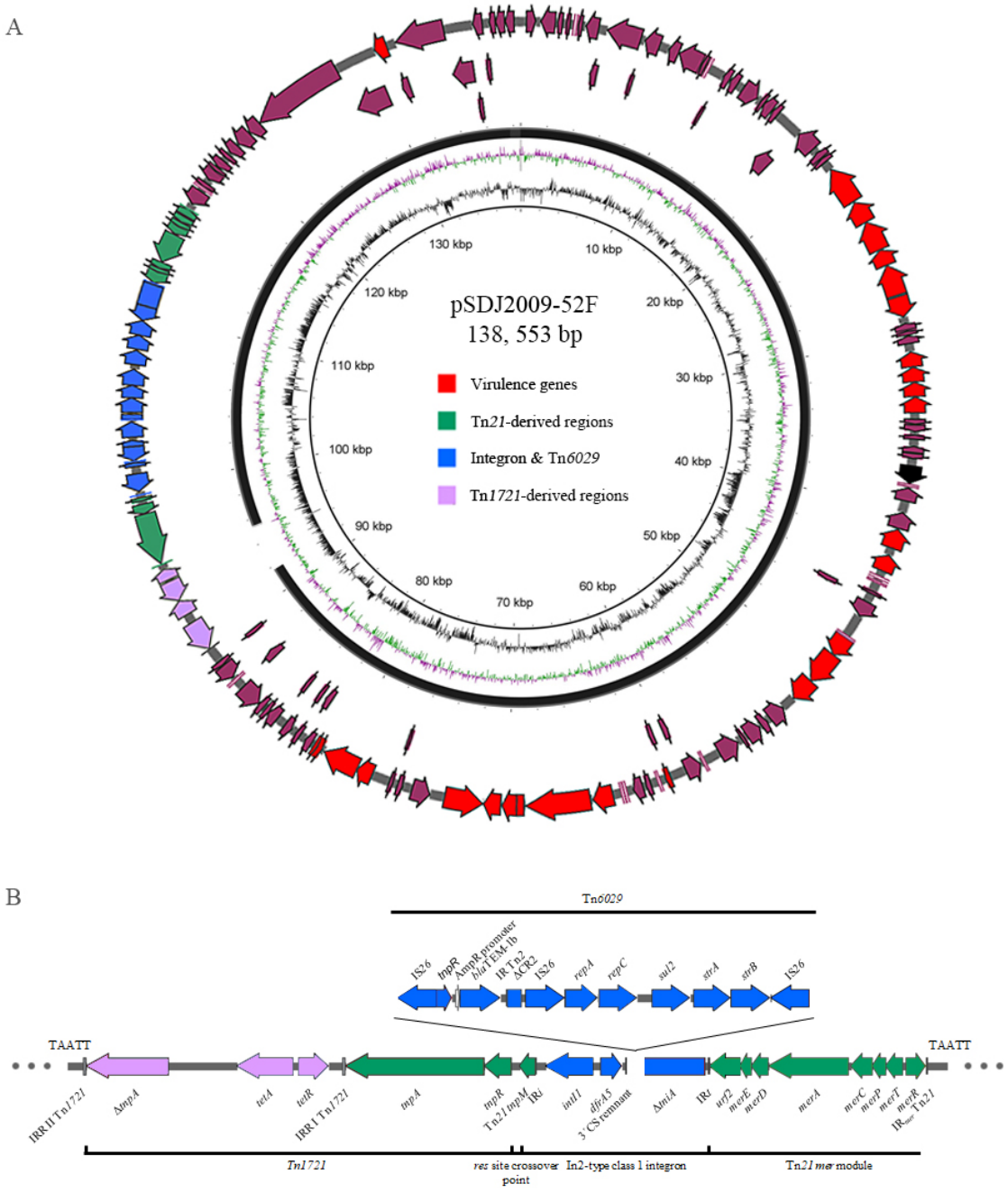


Figure 2: (A) Sequence of plasmid pSDJ2009-52F compared with pSF-088-1 using BLAST Ring Image Generator (BRIG). The green/purple histogram is the GC skew, the black histogram is the GC content and the black bar is the BLAST match with pSF-088-1. The outer ring is a schematic of pSDJ2009-52F created in SnapGene. Virulence-associated genes and the complex resistance locus/loci (CRL) are colour-coded. (B) CRL found in sequence type 58 (ST58) *Escherichia coli* strains 2009-49 and 2009-52. Genes are colour-coded to coincide with those in Fig. 2A.

Table 1: Putative virulence factors in ST58 strains 2009-49 and 2009-52.

Gene Function		Gene Name	% Identity	Location
Metal Resistance, Transport and Synthesis	Yersiniabactin receptor	*chuA*	98	chromosome
	Control of copper homeostasis	*copA*	98	chromosome
	multicopper oxidase	*cueO*	97	chromosome
	Transcriptional regulator	*cueR*	100	chromosome
	Copper and silver efflux membrane component	*cusA*	99	chromosome
	Copper and silver efflux membrane fusion protein	*cusB*	99	chromosome
	Copper and silver efflux outer membrane protein	*cusC*	98	chromosome
	Periplasmic copper binding protein	*cusF*	98	chromosome
	Response regulator	*cusR*	97	chromosome
	Histidine kinase	*cusS*	97	chromosome
	Putative type 1 secretion membrane-fusion protein	*etsA*	99	pSDJ2009-52F
	Putative type 1 secretion ATP binding protein	*etsB*	99	pSDJ2009-52F
	Putative type 1 secretion outer membrane protein	*etsC*	100	pSDJ2009-52F
	Pesticin/Yersiniabactin receptor	*fyuA*	99	chromosome
	Iron-related glycosyltransferase	*iroB*	100	pSDJ2009-52F
	Putative ABD transporter	*iroC*	99	pSDJ2009-52F
	Ferric enterochelin esterase	*iroD*	99	pSDJ2009-52F
	Siderophore esterase	*iroE*	99	pSDJ2009-52F
	Iron outer membrane receptor	*iroN*	100	pSDJ2009-52F
	Yersiniabactin biosynthesis	*irp1*	99	chromosome
	Yersiniabactin biosynthesis	*irp2*	99	chromosome
	Aerobactin biosynthesis	*iucA*	100	pSDJ2009-52F
	Aerobactin biosynthesis	*iucB*	100	pSDJ2009-52F
	Aerobactin biosynthesis	*iucC*	100	pSDJ2009-52F
	Aerobactin biosynthesis	*iucD*	99	pSDJ2009-52F
	Ferric aerobactin receptor	*iutA*	100	pSDJ2009-52F
	Iron/manganese transport system periplasmic binding protein	*sitA*	99	pSDJ2009-52F
	Iron/manganese transport system inner membrane protein	*sitB*	100	pSDJ2009-52F
	Iron/manganese transport system inner membrane protein	*sitC*	99	pSDJ2009-52F
	Iron/manganese transport system inner membrane protein	*sitD*	100	pSDJ2009-52F
Colicins and	Colicin V secretion protein	*cvaA*	100	pSDJ2009-52F
	Colicin V secretion protein	*cvaB*	99	pSDJ2009-52F

Colicin Immunity	Colicin V precursor	*cvaC*	100	pSDJ2009-52F
	Colicin V immunity protein	*cvi*	100	pSDJ2009-52F
Adhesion	Phase-variable biofilm formation autotransporter	*agn43*	98	chromosome
	M-agglutinin subunit	*bmaE*	100	chromosome
	Curli secretion channel	*csgG*	98	chromosome
	Tyrosine recombinase	*fimB*	99	chromosome
	Minor fimbrial subunit	*fimH*	99	chromosome
	G-fimbriae	*gafD*	99	chromosome
	putative fimbrial-like adhesin protein	*yfcV*	98	chromosome
Blood-associated	Auto-aggregating adhesin and invasion	*hek*	98	chromosome
	Haemolysin/cytolysin	*hlyE*	99	chromosome
	Avian haemolysin and regulator of outer membrane vesicle biogenesis	*hlyF*	100	pSDJ2009-52F
	Increased serum survival and resistance to complement	*iss*	100	pSDJ2009-52F
Other	Outer membrane protease, regulates the biogenesis of outer membrane vesicles	*ompT*	100	pSDJ2009-52F
	putative membrane transport protein	*shiF*	99	pSDJ2009-52F
	Conjugal transfer surface exclusion protein	*traT*	99	pSDJ2009-52F

References

1. Pitout JD. Extraintestinal Pathogenic Escherichia coli: A Combination of Virulence with Antibiotic Resistance. *Front Microbiol* 2012; **3**: 9.
2. Manges AR. Escherichia coli and urinary tract infections: the role of poultry-meat. *Clin Microbiol Infect* 2016; **22**: 122-9.
3. Nordstrom L, Liu CM, Price LB. Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. *Front Microbiol* 2013; **4**: 29.
4. Vincent C, Boerlin P, Daignault D et al. Food reservoir for Escherichia coli causing urinary tract infections. *Emerg Infect Dis* 2010; **16**: 88-95.
5. Dawes FE, Kuzevski A, Bettelheim KA 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**: e12754.
6. Reid CJ, Wyrsch ER, Roy Chowdhury P et al. Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microbial genomics* 2017; **3**.
7. Cain AK, Liu X, Djordjevic SP et al. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant Salmonella enterica serovar Typhimurium from Australian animals. *Microb Drug Resist* 2010; **16**: 197-202.
8. Levings RS, Lightfoot D, Partridge SR et al. The genomic island SGI1, containing the multiple antibiotic resistance region of Salmonella enterica serovar Typhimurium DT104 or variants of it, is widely distributed in other S. enterica serovars. *J Bacteriol* 2005; **187**: 4401-9.
9. Reid C, Roy Chowdhury P, Djordjevic SP. *Tn6026 and Tn6029 are found in complex resistance regions mobilized by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae*, 2015.
10. 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**: e0115781.
11. Saenz Y, Vinue L, Ruiz E et al. Class 1 integrons lacking qacEDelta1 and sul1 genes in Escherichia coli isolates of food, animal and human origins. *Vet Microbiol* 2010; **144**: 493-7.
12. Venturini C, Beatson SA, Djordjevic SP et al. Multiple Antibiotic Resistance Gene Recruitment onto the Enterohemorrhagic Escherichia coli Virulence Plasmid. *FASEB J* 2010; **24**: 1160-6.
13. Venturini C, Hassan KA, Roy Chowdhury P 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**: e78862.
14. Mangat CS, Bekal S, Irwin RJ et al. A Novel Hybrid Plasmid Carrying Multiple Antimicrobial Resistance and Virulence Genes in Salmonella enterica Serovar Dublin. *Antimicrobial Agents and Chemotherapy* 2017; **61**: e02601-16.
15. Porse A, Schønning K, Munck C et al. Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. *Molecular biology and evolution* 2016; **33**: 2860-73.
16. Darling AE, McKinnon J, Worden P et al. A draft genome of Escherichia coli sequence type 127 strain 2009-46. *Gut pathogens* 2014; **6**: 32.

17. Treangen TJ, Ondov BD, Koren S et al. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology* 2014; **15**: 524.
18. Abraham S, Jordan D, Wong HS et al. First detection of extended-spectrum cephalosporin- and fluoroquinolone-resistant Escherichia coli in Australian food-producing animals. *Journal of global antimicrobial resistance* 2015; **3**: 273-7.
19. Bettelheim KA, Hornitzky MA, Djordjevic SP et al. Antibiotic resistance among verocytotoxigenic Escherichia coli (VTEC) and non-VTEC isolated from domestic animals and humans. *J Med Microbiol* 2003; **52**: 155-62.

Chapter 6: Molecular analysis of an IncF ColV-like plasmid lineage that carries a complex resistance locus with a trackable genetic signature

Molecular analysis of an IncF ColV-like plasmid lineage that carries a complex resistance locus with a trackable genetic signature

Jessica McKinnon[1], Piklu Roy Chowdhury[1,2] and Steven P Djordjevic[1]

[1]ithree Institute, University of Technology Sydney, PO Box 123, Broadway, Sydney, NSW, Australia, 2007
[2]Department of Primary Industries, Elizabeth Macarthur Agriculture Institute, PMB 4008, Camden, NSW 2567, Australia

Abstract

IncF plasmids are an important plasmid incompatibility group that are currently restricted to the Enterobacteriaceae. The ColV subgroup of these plasmids carry an important repertoire of virulence-associated genes (VAGs) that contribute to the ability of avian pathogenic *E. coli* to cause disease in poultry. VAGs found on ColV plasmids have also been linked with urosepsis and meningitis in humans but the mechanisms that elicit these disease conditions are not well understood. Recently we described the sequence of a ColV plasmid pSDJ2009-52F that carried the typical repertoire of VAGs and a complex resistance gene locus flanked by IS*26*, an insertion element that plays an important role in mobilizing antibiotic resistance genes on plasmids and genomic islands. We recovered complete ColV-like plasmid sequences from public databases that shared more than 80% sequence identity with pSDJ2009-52F in geographically diverse regions of the world over a twenty year time frame. Previously we noted that pSDJ2009-52F carries a unique genetic signature in the class 1 integron within the complex resistance locus (CRL) that was presumably created by the action of IS*26*. Here we show that most ColV-like plasmids that are closely related to pSDJ2009-52F also carry the same signature. Our studies provide insight into how these signature-bearing plasmids and the mobile genetic elements they carry traffic between *E. coli* sequence types (ST) over large geographic distances.

Keywords: ExPEC; virulence plasmid; complex resistance region; Tn*6029*; ColV

Introduction:

Escherichia coli that have colonised tissue sites removed from the gastrointestinal tract are referred to as extraintestinal pathogenic *E. coli* (ExPEC). There are three subdivisions of this classification; uropathogenic *E. coli* (UPEC), neonatal meningitis associated *E. coli* (NMEC) and avian pathogenic *E. coli* (APEC) however the range of *E. coli* causing extra-intestinal infection is incredibly diverse[1]. ExPEC cause disease in humans, agriculturally important animals and

companion animals and are often exposed to antimicrobial selection pressures. Knowledge of mobile genetic elements carried by ExPEC is critical to understanding their role in pathogenesis and in the capture and spread of drug resistance.

Plasmids are important vehicles for capturing and assembling antibiotic resistance, heavy metal resistance and virulence genes[2-5] and their presence can impact how pathogens evolve[2, 6-8]. Numerous plasmid incompatibility types circulate in human-sourced ExPEC but IncF plasmids play a key role in housing genes conferring resistance to last line antibiotics, including the New Delhi metallo-β-lactamase gene *bla*$_{\text{NDM-1}}$[9] the cephalosporinase gene *bla*$_{\text{CTX-M}}$[10] and various virulence genes[8]. ColV plasmids range between 80 and 180 kb and often carry an extensive suite of virulence determinants, including, but not limited to, increased serum survival (*iss*) and aerobactin biosynthesis (*iuc* operon) and other iron acquisition genes[11]. IncF ColV plasmids are widespread among APEC strains and can be found in commensal human flora[12] and in *E. coli* causing urosepsis[8, 13, 14] and neonatal meningitis[15, 16].

IS*26* is an insertion element that plays a major role in the capture and assembly of multiple antibiotic resistance genes in complex resistance gene regions found on diverse plasmid backbones[3, 17-19]. IS*26* has no recognized target site specificity[18] nonetheless insertions next to class 1 are reported frequently[20, 21] possibly because they are unlikely to infer a significant fitness cost and afford the host bacterium more than one mechanism to acquire antibiotic resistance genes. IS*26* often creates random deletions in DNA that abut its insertion site. We have previously investigated IS*26*-associated deletion events as unique genetic signatures for the purpose of tracking genetic elements that purvey multiple drug resistance[4, 20, 22]. Collectively these and more recent genomic surveillance studies have shown that the conserved structure of class 1 integrons is frequently altered by the action of insertion elements, particularly IS*26*[8, 21-24]. One such class 1 integron structure contains the trimethoprim resistance gene cassette *dfrA5* and only 24 bp of the 3´-CS. It has been reported in association with multiple drug resistant *E. coli* from the faeces of Australian production animals[4, 20, 21] and in human patients with haemorrhagic colitis[3] and urosepsis[8] and is often carried on Tn*3* family, mercury resistance transposons.

Recently we characterised a ColV-like plasmid, pSDJ2009-52F, in a multiple drug resistant *E. coli* with ST58 from a patient with urosepsis. Although ST58 *E. coli* are found in healthy and diseased agriculturally important animals[24-26] and flies[27], the environment[28], and wildlife[29], it is

not noted to be a frequent cause of human disease[8]. Plasmid acquisition is known to play a critical role in pathogen evolution[2, 6, 7]. Carriage of pSDJ2009-52F may have played an important role in the ability of ST58 strain 2009-49 (urine isolate) to cause urinary tract infection followed by urosepsis[8] (2009-52). Here we searched public databases for plasmids closely related to pSDJ2009-52F. We identified IncF ColV plasmids similar to pSDJ2009-52F from different countries, years, and hosts and from *E. coli* with diverse sequence types and investigated their phylogeny.

Methods:

Sequencing of pSDJ2009-52F

Methods used to isolate, sequence and characterize strain 2009-52 and the plasmids it carries, including pSDJ2009-52F have been described [8].

Single Nucleotide Polymorphism (SNP) Analysis

Parsnp from the Harvest Tools suite[30] was used to identify SNPs between pSDJ2009-52 and the five most closely related plasmids in GenBank as determined by BLASTn analysis. The x flag was implemented to filter out recombination events, resulting in an average alignment of 62% between the plasmids. Gingr, also from Harvest Tools, was used to visualise the SNP locations in the .ggr file and the matching co–ordinates confirmed from the exported .vcf file.

BEAST2 Analysis

To identify plasmid evolution over time we used a Bayesian Markov chain Monte Carlo method implemented in the BEAST2 package[31]. The GTR-I substitution model was used as the plasmids were very closely related. A strict molecular clock on a constant population size coalescent model was assumed and analysed as reported previously[32].

Results:

Plasmid Information

Plasmid pSDJ2009-52F was used as the query for BLASTn analysis to determine the most closely related plasmids in GenBank. Twelve plasmids with high sequence identity (>80%) to pSDJ2009-52F are shown in Table 1 along with a description of their salient features. Notably, eight of eleven plasmids were isolated from *E. coli* sourced from humans including strains from patients with gastrointestinal disease and strains from apparently healthy humans (plasmids cited as site: faeces excluding strain 2013C-4390). These plasmids belong to *E. coli* from eight different STs and were from geographically diverse regions around the globe (Table 1) and from multiple sources including humans, chicken, canine, swine and turkey. The strains carrying these related plasmids were sourced over a 22 year period with 1990 being the earliest isolation date. Besides pAPEC-O78-ColV (IncF24) all plasmids were plasmid MLST F2:A-:B1.

Virulence and Antibiotic Resistance Gene Carriage

A full set of genes typical of ColV plasmids[11] were observed in all but two plasmids, p2013-4390 and 134q. These include the *iro* operon (iron-related), *iuc* operon for aerobactin biosynthesis, ferric aerobactin receptor *iutA*, *sit* operon for iron/manganese transport, *cvaABC*/*cvi* for colicin V secretion and immunity and colicin Ia, putative type 1 secretion system genes *etsABC*, avian haemolysin *hlyF*, outer membrane protease *ompT*, conjugal transfer exclusion protein *traT* and *iss* for increased serum survival and complement resistance. Plasmid 134q was missing the *etsABC* genes as well as *hlyF* and *ompT*, and plasmid 'unnamed' from strain 2013C-4390 was missing the *iutA* gene and *iuc* operon. Due to the presence of an IS*1* immediately following the sit operon in this plasmid, the normally adjacent *iuc* operon is likely the target of an IS*1* mediated deletion. The remaining nine plasmids contained the full set of standard IncF ColV virulence genes present in pSDJ2009-52F albeit with occasional single nucleotide polymorphisms (SNPs) (Table 2).

As expected, antibiotic resistance genes were less conserved across the plasmids with only one plasmid, pECAZ147_1, exhibiting an identical profile as pSDJ2009-52F. Nine of 11 plasmids contained a *bla*$_{\text{TEM-1}}$ ampicillin resistance gene variant, 8/11 contained the *sul2-strA-strB*

sulphonamide and streptomycin resistance gene cluster, 9/11 contained trimethoprim resistance gene *dfrA5* and 6/11 contained tetracycline resistance gene *tetA*, while sulphonamide resistance gene *sul1* and kanamycin resistance gene *aphA1* were singularly represented. *dfrA5* is a resistance gene cassette typically linked with carriage of a class 1 integron. As such the low carriage rate of *sul1* is unusual given the sulphonamide resistance gene is typically resident in the 3′-CS of class 1 integrons.

Comparative SNP Analyses

Plasmids that shared at least 90% sequence identity over the full length of plasmid pSDJ2009-52F, specifically pECAZ147_1, plasmid 'unnamed' from strain 2013C-4390, pG749_1, pDB4277, pSF-088-1 and pCERC4, were examined further. The high similarity of the plasmid backbones is evident in Figure 1. As described above, the major differences among these plasmids reside in the assembly of the CRL (complex resistance locus) resident in each plasmid. pSDJ2009-52F contains an atypical class 1 integron containing a *dfrA5* trimethoprim resistance cassette bearing only 24 bp of the 3′- CS carried on a hybrid Tn*1721*/Tn*21* transposon (Supplementary Figure 1 A). A copy of the composite transposon Tn*6029* which harbours the $bla_{\text{TEM-1}}$ gene and *sul2-strA-strB* gene cluster encoding resistance to ampicillin, streptomycin and sulphonamides abuts the truncated 3′- CS. The resistance gene locus is essentially identical with the locus found on EHEC plasmid pO26-CRL_{111} sourced from a case of haemorrhagic colitis in 1998[3, 4], however the left arm of the transposon lacks the *aphA1* gene of Tn*6026* (Tn*6029*) and the mercury resistance transposon on pO26-CRL111 lacks a copy of Tn*1721*. pECAZ147_1 and 'unnamed' (strain 2013C-4390) contain an identical resistance region comprising the class 1 integron, hybrid Tn*21*/*1721* transposon and Tn*6029* to pSDJ2009-52F, whereas pG749_1, pDB4277 and pSF-088-1 are missing the leftmost arm of Tn*1721* including the tetracycline resistance genes *tetA* and *tetR*. pCERC4 was isolated from an *E. coli* from a healthy human living in Sydney and lacks Tn*6029* (Supplementary Figure 1 B).

Parsnp identified 20 unique SNPs across all plasmid backbones as compared to pSDJ2009-52 (62% sequence coverage shared across all plasmids), two of which were shared across all six plasmids. These two SNPs were located in the putative type 1 secretion system gene *etsA* gene and conjugal transfer gene *traD.* Neither of these SNPs were predicted to affect protein function using the PROVEAN web program[33] (http://provean.jcvi.org/seq_submit.php). pECAZ147_1

contained two additional SNPs in *traD,* however these were also tolerated. The closest relative to pSDJ2009-52F was pDB4277. It differed from pSDJ2009-52F only by the two SNPs described above. pDB4277 was isolated from a uropathogenic *E. coli* strain of unknown sequence type in the United Kingdom in 2004. The unnamed plasmid from Shiga-toxin producing *E. coli* strain 2013C-4390 also only contained these 2 SNPs as well as an identical atypical In22 class 1 integron and hybrid Tn*1721*/Tn*21* transposon to pSDJ2009-52F but lacked the *iutA* gene and *iuc* operon. pCERC4 and pG749_1 were the next most closely related with 4 SNPs each, two of which were unique in both cases. These were located in an uncharacterized open reading frame and the relaxase gene *traI* involved in plasmid conjugation for pCERC4, and mercury resistance protein *merP* and *traI* for pG479_1. Plasmid pECAZ147_1 from an ST540 *E. coli* isolated in Israel in 2012 housed an identical complex antimicrobial resistance region to pSDJ2009-52F but contained 4 unique SNPs. These were in conjugative transfer proteins *traD* (two SNPs) and *traG,* and an uncharacterised open reading frame. The most disparate backbone sequence was pSF-088-1, with 10 unique SNPs, eight of which were within mobile elements IS*1* and IS*2*. The remaining two were within uncharacterized open reading frames.

BEAST2 Analysis

Time-measured phylogeny (Supplementary Figure 2) estimated the oldest common ancestor from which these plasmids have evolved is 4.6 thousand years. Interestingly, plasmid 'unnamed 1' from strain FAP1 does not appear related to the others in terms of evolution despite showing 89% sequence identity to pSDJ2009-52F and a near identical virulence gene profile. The closest evolutionary relation to pSDJ2009-52F is plasmid 'unnamed' from strain 2013C-4390 isolated from a Shiga-toxin producing *E. coli*.

Molecular Signatures

Plasmid pSDJ2009-52F houses a characteristic IS*26*-mediated deletion in the class 1 integron and retains only 24 base pairs of the 3′-CS. Notably, most plasmids shown in Table 1 carried this signature, and five of these fell above the 90% query cover cutoff (Figure 1).

Discussion

IncF plasmids are a prominent incompatibility group amongst *E. coli* from human and animal sources[34]. Their dominance is likely due to their conjugative nature in conjunction with their host range being limited to Enterobacteriaceae. Consistent with this view we observed plasmids that share high sequence identity with pSDJ2009-52F in *E. coli* with diverse sequence types (Table 1). Of the 11 strains, 8 unique sequence types were represented including the multi-resistant clonal group ST131 (pG749_1: H22, pCERC5: undisclosed), well documented for causing ESBL producing and fluoroquinolone resistant extra-intestinal infection[35]. IncF plasmids described here not only contain extensive virulence gene cargo associated with extraintestinal disease, but nine of eleven plasmids contain a class 1 integron, a reliable proxy for multiple drug resistance[36, 37]. Consistent with this, all nine plasmids carry numerous antibiotic resistance genes. IncF type plasmids are inherently conjugative if the *tra* locus is intact. The observation that diverse *E. coli* carry MDR IncF ColV plasmids is a cause for concern, as these strains tend to be missed due to the current research bias towards those that have established associations with genes conferring resistance to extended spectrum β lactams[34].

Two plasmids in this study, pCERC4 and pSF-088-1, were isolated from strains of *E. coli* ST95, a pandemic lineage that has not yet received the 'elite' global status of ST131 due to the relatively low frequency of multi-resistant strains[38]. However, ST95 is responsible for a significant number of human ExPEC infections as well as systemic disease in poultry[39-42]. One study[42] found that a cohort of 116 ST95 strains form biofilms, resist host serum bactericidal activity, and adhere to and subsequently invade mammalian kidney cell lines. They found no distinguishable differences between strains of human and avian origin in terms of virulence attributes and pathogenesis. This poses a serious risk in terms of human clinical infection as ST95 strains are generally virulent and recognised human pathogens[38]. Evidence linking APEC with episodes of human disease is mounting[41, 43-46]. Carriage of ColV plasmids likely plays a key role as IncF ColV plasmids are typically associated with APEC[11]. The report of isolation of MDR ColV plasmid pCERC4 from commensal faecal *E. coli* of a healthy human that had not received antibiotics for at least six months[47] shows that plasmids carrying important virulence gene cargo reside in the flora of healthy humans and pose a potential threat for UTI[12, 48].

The highly related plasmids represented here have been isolated from three animal sources as well as humans and were derived from diverse geographic locations (Table 1). The similarity in virulence gene content and the minor variations in complex antimicrobial resistance loci amongst the IncF plasmids described here highlights the global dissemination of the IncF incompatibility group and their propensity to house pathogenicity determinants. ColV plasmids are important in this regard because they are resident in APEC that cause disease in poultry, carry a significant armoury of VAGs (virulence associated genes) associated with human and food animal disease, and acquire antimicrobial resistance creating strains capable of high risk infection in a single genetic transfer event[7, 8]. Evident in the analysis described here is the propensity of ColV plasmids to acquire CRL. Owing to the numerous mobile elements involved in the assembly of these CRL, they can undergo rearrangement and recombination events in response to selective pressure exerted by the niche they inhabit. Time-measured phylogeny (Supplementary Figure S2) estimated the oldest common ancestor from which the plasmids in this study have evolved is 4.6 thousand years. Antimicrobial selective pressures in recent times have likely played a role in the formation and rearrangement of CRL described in this plasmid cohort but natural selection of the ColV plasmid backbones predates the modern antibiotic era. Therefore it is possible that this backbone has, in part, become established due to the beneficial cargo it carries which includes efflux proteins, colicin/colicin immunity, heavy metal resistance and increased serum survival.

The observation of this plasmid in several animal species as well as human commensal and pathogenic bacteria is cause for concern. Our studies underscore the need for a 'One Health' approach to understand the real impact of reliance on antimicrobials to control infectious disease and to produce food for a burgeoning world population. Future directions of this research may include conjugation studies of pSDJ2009-52F.

CRediT Statement

Jessica McKinnon: Conceptualisation, Data curation, Formal analysis, Investigation, Visualisation, Writing – original draft (lead). **Piklu Roy Chowdhury:** Investigation, Resources, Supervision, Visualisation, Writing – review and editing. **Steven P.**

Djordjevic: Conceptualisation, Funding acquisition, Project administration, Supervision (lead), Writing - original draft, Writing – review and editing.

Funding Statement

This work was supported by Ausgem, a collaborative partnership between the New South Wales Department of Primary Industries' Elizabeth Macarthur Agricultural Institute (EMAI) and the ithree Institute at the University of Technology Sydney (UTS). JM is the recipient of an Australian Government Research Training Program Scholarship.

Table 1: Plasmids with high sequence identity to pSDJ2009-52F.

Species	Strain Name	Plasmid Name	Percent identity	Accession	Year	Host Strain ST	pMLST	Source	Site	Location
E.coli	Ecol AZ147	pECAZ147_1	100	CP018994.1	2012	540	F2:A-:B1	human		Israel
	G749	pG749_1	97	CP014489.1	2010	131	F2:A-:B1	human		Seattle
	DB04277	pDB4277	97	KP398867.1	2004		F2:A-:B1	human	urine	United Kingdom
	SF-088	pSF-088-1	97	CP012636.1	2007	95	F2:A-:B1	human	blood	San Francisco
	1.10-R8	pCERC4	92	KU578032.1	2012	95	F2:A-:B1	human	faeces	Australia
	2013C-4390	unnamed	91	CP027485.1	2013	675	F2:A-:B1	human	faeces	
	FAP1	unnamed 1	89	CP009579.1	2011	453	F-:A-:B1	pig	faeces	Netherlands
	11.3-R3	pCERC5	88	KU664810.1	2010	131	F2:A-:B1	human	faeces	Australia
	144	134q	87	CP023363.1	2002	38	F2:A-:B1	canine		Edinburgh
	O78-789	pAPEC-O78-ColV	86	CP010316.1	1990	88	F24:A-:B1	turkey	blood	Israel
	14.3-R4	pCERC9	86	KY007017.1	2010	2705	F2:A-:B1	human	faeces	Australia
	2009-52	**pSDJ2009-52F**	**ref**		**2009**	**58**	**F2:A-:B1**	**human**	**blood**	**Sydney**
Salmonella	CVM29188	pCVM29188_146	80	CP001122.1	2003		F24:A-:B1	chicken breast		Georgia
	CS0010A	pCS0010A	80	CP002090.1			F24:A-:B1	chicken	cloaca	

Cells are blank if information was not provided or could not be ascertained.

Plasmid pSDJ2009-52F[8] is identified in bold text.

Table 2: Percentage sequence identity across whole open reading frame based on BLASTn analyses to virulence genes found on IncF colV plasmids.

Plasmid Name	*iutA*	*iroN*	*iucD*	*sitD*	*cvaA*	*cvaB*	*cvaC*	*cvi*	*etsA*	*etsB*	*etsC*	*hlyF*	*ompT*	*traT*	*iss*
pSDJ2009-52F	**100**	**100**	**99.92**	**100**	**100**	**99.91**	**100**	**100**	**99.92**	**99.9**	**100**	**100**	**100**	**98.91**	**100**
pECAZ147_1	100	100	99.92	100	100	99.91	100	100	100	99.9	100	100	100	98.91	100
pG749_1	100	100	99.92	100	100	99.91	100	100	100	99.9	100	100	100	98.91	100
pSF-088-1	100	100	99.84	100	100	99.91	100	100	100	99.9	100	100	100	98.91	100
pDB4277	100	100	99.92	100	100	99.91	100	100	100	99.9	100	100	100	98.91	100
pCERC4	100	100	99.92	100	100	99.91	100	100	100	99.9	100	100	100	98.91	100
p2013C-4390		100		100	100	99.91	100	100	100	99.85	100	100	100	98.91	100
pFAP1	100	100	99.92	100	100	99.86	100	100	100	99.9	100	100	100	99.73	100
pCERC5	100	100	99.92	100	100	99.91	100	100	100	99.9	100	100	100	98.91	100
134q	100	100	99.92	100	100	99.86	100	100						98.91	100
pCERC9	99.86	100	99.92	100	100	99.91	100	100	100	99.85	99.93	100	100	98.91	100
pAPEC-O78-ColV	100	100	99.92	100	100	99.91	100	100	100	99.9	100	100	100	98.63	100

Representative genes have been used as markers for larger operons (eg. *iroN*, *iucD*, *sitD*). Italicised plasmids were those not used in comparative analyses. Blank cells indicate gene absence.

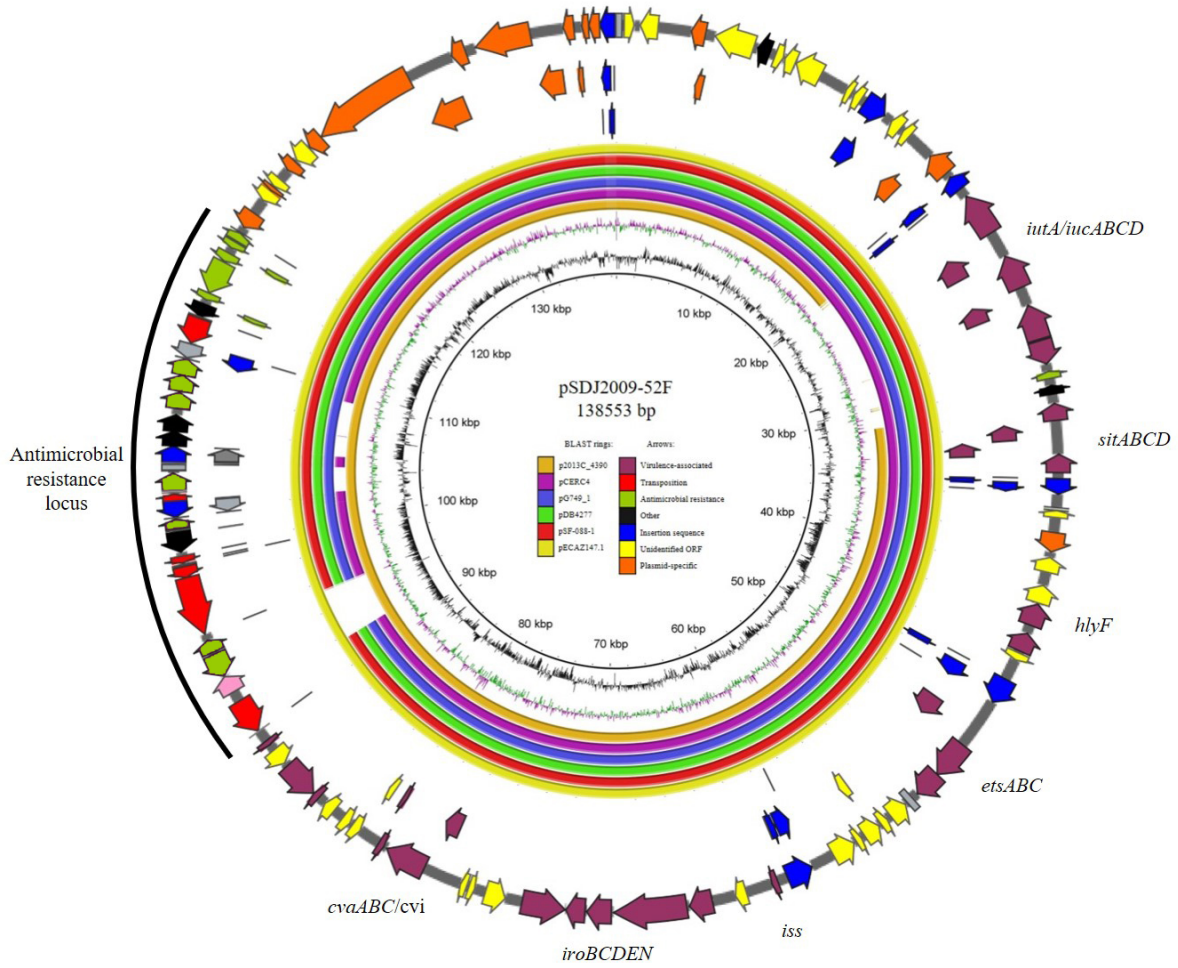
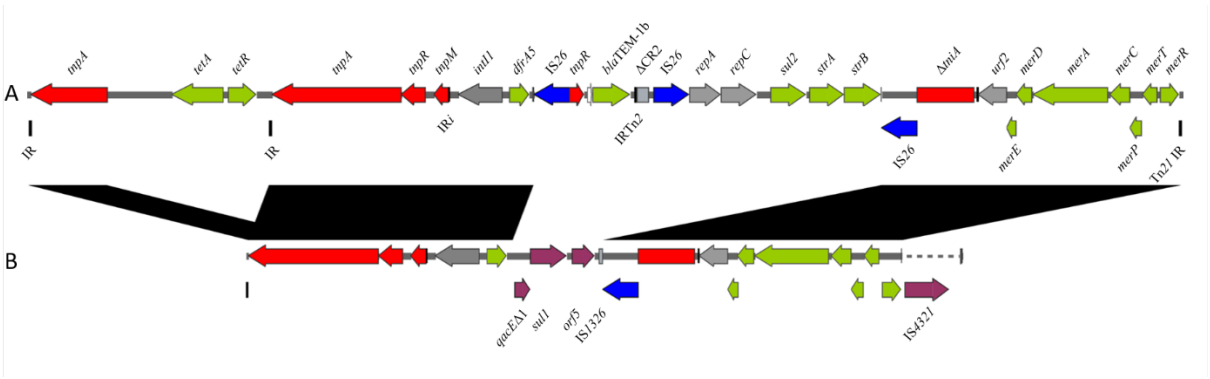
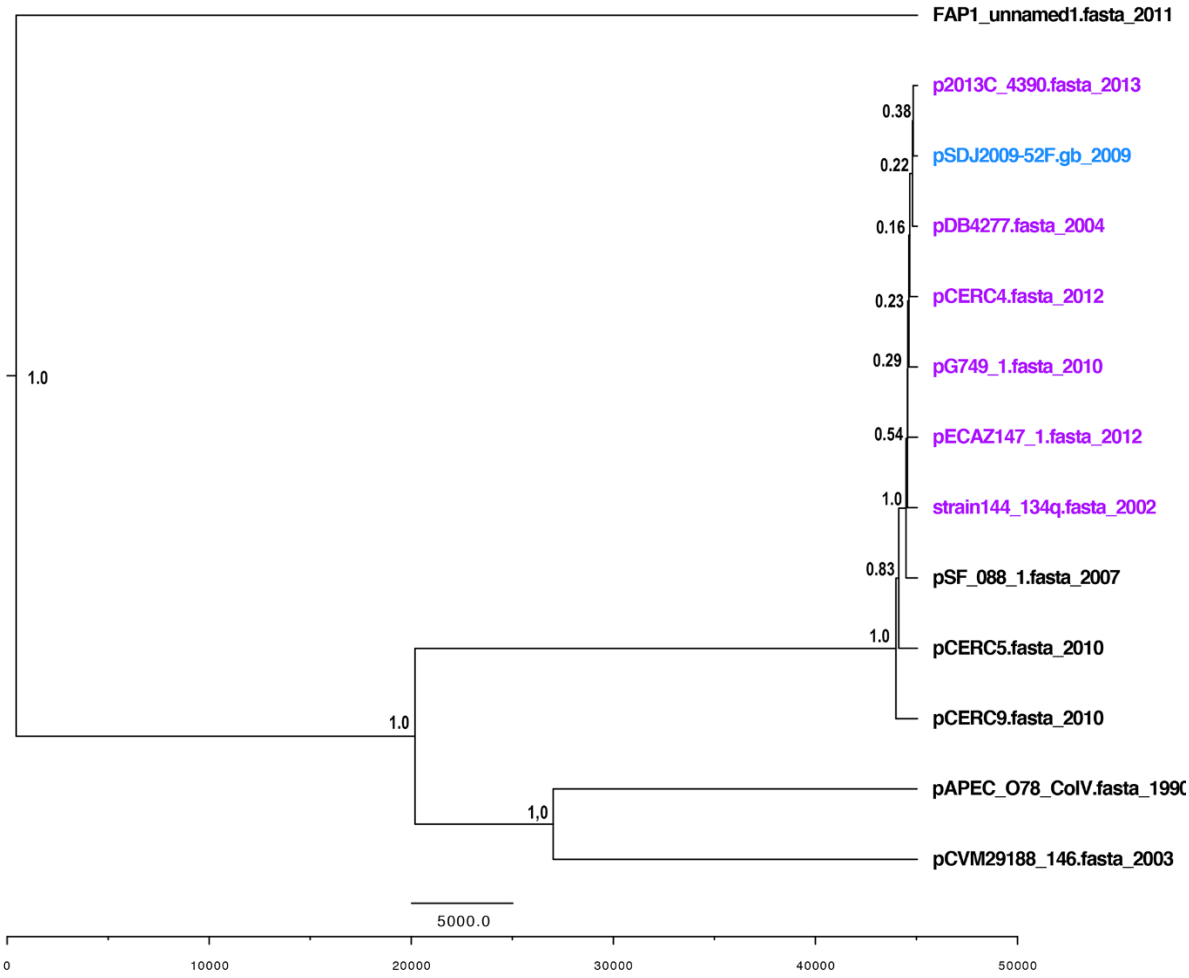


Figure 1: BLAST comparison, produced by BRIG, of plasmids in GenBank matching at least 90% sequence identity to pSDJ2009-52F via BLAST analysis. Outside ring (arrows) shows gene content for pSDJ2009-52F in correlation with BLAST results.



Supplementary Figure 1: Comparison between antimicrobial resistance gene regions found in pSDJ2009-52F (A) and pCERC4 (B). Easyfig[49] and SnapGene were used to generate the figure. Antimicrobial resistance genes are represented by green arrows, transposition related genes by red arrows and insertion elements by blue arrows.



Supplementary Figure 2: BEAST2 tree of plasmid sequences included in this analysis with disclosed years of isolation. pSDJ2009-52F is shown in blue and those in violet are most closely related. Scale bar indicates age in years.

References

1. Vila J, Sáez-López E, Johnson JR et al. Escherichia coli: an old friend with new tidings. *FEMS Microbiology Reviews* 2016; **40**: 437-63.
2. Fang LX, Li XP, Deng GH et al. High Genetic Plasticity in Multidrug-Resistant Sequence Type 3-IncHI2 Plasmids Revealed by Sequence Comparison and Phylogenetic Analysis. *Antimicrob Agents Chemother* 2018; **62**.
3. Venturini C, Beatson SA, Djordjevic SP et al. Multiple Antibiotic Resistance Gene Recruitment onto the Enterohemorrhagic Escherichia coli Virulence Plasmid. *FASEB J* 2010; **24**: 1160-6.
4. Venturini C, Hassan KA, Roy Chowdhury P 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**: e78862.
5. Wang X, Hao H, Xu Z et al. Plasmid-mediated multidrug resistance and virulence in an avian pathogenic Escherichia coli strain isolated in China. *Journal of global antimicrobial resistance* 2014; **2**: 57-8.
6. Jackson RW, Vinatzer B, Arnold DL et al. The influence of the accessory genome on bacterial pathogen evolution. *Mobile genetic elements* 2011; **1**: 55-65.
7. Johnson TJ, Thorsness JL, Anderson CP et al. Horizontal gene transfer of a ColV plasmid has resulted in a dominant avian clonal type of Salmonella enterica serovar Kentucky. *PLoS One* 2010; **5**: e15524.
8. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. *Int J Antimicrob Agents* 2018; **52**: 430-5.
9. Zheng B, Yu X, Xu H et al. Complete genome sequencing and genomic characterization of two Escherichia coli strains co-producing MCR-1 and NDM-1 from bloodstream infection. *Scientific reports* 2017; **7**: 17885.
10. Kawamura K, Sugawara T, Matsuo N et al. Spread of CTX-Type Extended-Spectrum beta-Lactamase-Producing Escherichia coli Isolates of Epidemic Clone B2-O25-ST131 Among Dogs and Cats in Japan. *Microb Drug Resist* 2017; **23**: 1059-66.
11. Johnson TJ, Siek KE, Johnson SJ et al. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian Escherichia coli strains. *J Bacteriol* 2006; **188**: 745-58.
12. 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; **84-85**: 11-9.
13. Fernandez-Beros ME, Kissel V, Lior H et al. Virulence-related genes in ColV plasmids of Escherichia coli isolated from human blood and intestines. *J Clin Microbiol* 1990; **28**: 742-6.
14. Peigne C, Bidet P, Mahjoub-Messai F 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**: 2272-84.
15. Lemaitre C, Bidet P, Benoist JF et al. The ssbL gene harbored by the ColV plasmid of an Escherichia coli neonatal meningitis strain is an auxiliary virulence factor boosting the production of siderophores through the shikimate pathway. *J Bacteriol* 2014; **196**: 1343-9.

16. Zhu Ge X, Jiang J, Pan Z 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**: e112048.
17. Cain AK, Liu X, Djordjevic SP et al. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant Salmonella enterica serovar Typhimurium from Australian animals. *Microb Drug Resist* 2010; **16**: 197-202.
18. Harmer CJ, Moran RA, Hall RM. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* 2014; **5**: e01801-14.
19. Yau S, Liu X, Djordjevic SP et al. RSF1010-like plasmids in Australian Salmonella enterica serovar Typhimurium and origin of their sul2-strA-strB antibiotic resistance gene cluster. *Microb Drug Resist* 2010; **16**: 249-52.
20. Dawes FE, Kuzevski A, Bettelheim KA 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**: e12754.
21. Reid CJ, Wyrsch ER, Roy Chowdhury P et al. Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microbial genomics* 2017; **3**.
22. 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**: e0115781.
23. Roy Chowdhury P, McKinnon J, Liu M et al. Multidrug Resistant Uropathogenic Escherichia coli ST405 With a Novel, Composite IS26 Transposon in a Unique Chromosomal Location. *Frontiers in Microbiology* 2019; **9**.
24. Cummins ML, Reid C, Roy Chowdhury P et al. Whole genome sequence analysis of Australian avian pathogenic E. coli that carry the class 1 integrase gene. *MGen* 2019.
25. Dahmen S, Metayer V, Gay E et al. Characterization of extended-spectrum beta-lactamase (ESBL)-carrying plasmids and clones of Enterobacteriaceae causing cattle mastitis in France. *Vet Microbiol* 2013; **162**: 793-9.
26. Wu G, Ehricht R, Mafura M et al. Escherichia coli isolates from extraintestinal organs of livestock animals harbour diverse virulence genes and belong to multiple genetic lineages. *Vet Microbiol* 2012; **160**: 197-206.
27. Blaak H, Hamidjaja RA, van Hoek AH et al. Detection of extended-spectrum beta-lactamase (ESBL)-producing Escherichia coli on flies at poultry farms. *Appl Environ Microbiol* 2014; **80**: 239-46.
28. Sacramento AG, Fernandes MR, Sellera FP et al. Genomic analysis of MCR-1 and CTX-M-8 co-producing Escherichia coli ST58 isolated from a polluted mangrove ecosystem in Brazil. *Journal of global antimicrobial resistance* 2018; **15**: 288-9.
29. Dolejska M, Masarikova M, Dobiasova H et al. High prevalence of Salmonella and IMP-4-producing Enterobacteriaceae in the silver gull on Five Islands, Australia. *J Antimicrob Chemother* 2016; **71**: 63-70.
30. Treangen TJ, Ondov BD, Koren S et al. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology* 2014; **15**: 524.
31. Bouckaert R, Heled J, Kuhnert D et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology* 2014; **10**: e1003537.

32. Duchene S, Holt KE, Weill FX et al. Genome-scale rates of evolutionary change in bacteria. *Microbial genomics* 2016; **2**: e000094.
33. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics (Oxford, England)* 2015; **31**: 2745-7.
34. Rozwandowicz M, Brouwer MSM, Fischer J et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* 2018; **73**: 1121-37.
35. Pitout JD, DeVinney R. Escherichia coli ST131: a multidrug-resistant clone primed for global domination. *F1000Research* 2017; **6**.
36. Gillings MR, Gaze WH, Pruden A et al. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* 2015; **9**: 1269-79.
37. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. *Microbiology and molecular biology reviews : MMBR* 1999; **63**: 507-22.
38. Riley LW. Pandemic lineages of extraintestinal pathogenic Escherichia coli. *Clin Microbiol Infect* 2014; **20**: 380-90.
39. Adams-Sapper S, Diep BA, Perdreau-Remington F et al. Clonal composition and community clustering of drug-susceptible and -resistant Escherichia coli isolates from bloodstream infections. *Antimicrob Agents Chemother* 2013; **57**: 490-7.
40. Heidemann Olsen R, Bisgaard M, Christensen JP et al. Pathology and Molecular Characterization of Escherichia Coli Associated With the Avian Salpingitis-Peritonitis Disease Syndrome. *Avian diseases* 2016; **60**: 1-7.
41. Mora A, Viso S, López C et al. Poultry as reservoir for extraintestinal pathogenic Escherichia coli O45:K1:H7-B2-ST95 in humans. *Veterinary Microbiology* 2013; **167**: 506-12.
42. Nandanwar N, Janssen T, Kuhl M et al. Extraintestinal pathogenic Escherichia coli (ExPEC) of human and avian origin belonging to sequence type complex 95 (STC95) portray indistinguishable virulence features. *Int J Med Microbiol* 2014; **304**: 835-42.
43. Danzeisen JL, Wannemuehler Y, Nolan LK et al. Comparison of multilocus sequence analysis and virulence genotyping of Escherichia coli from live birds, retail poultry meat, and human extraintestinal infection. *Avian diseases* 2013; **57**: 104-8.
44. Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J et al. Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clin Microbiol Infect* 2011; **17**: 873-80.
45. Liu CM, Stegger M, Aziz M et al. Escherichia coli ST131-H22 as a Foodborne Uropathogen. *MBio* 2018; **9**.
46. Maluta RP, Logue CM, Casas MR et al. Overlapped sequence types (STs) and serogroups of avian pathogenic (APEC) and human extra-intestinal pathogenic (ExPEC) Escherichia coli isolated in Brazil. *PLoS One* 2014; **9**: e105016.
47. Moran RA, Hall RM. Analysis of pCERC7, a small antibiotic resistance plasmid from a commensal ST131 Escherichia coli, defines a diverse group of plasmids that include various segments adjacent to a multimer resolution site and encode the same NikA relaxase accessory protein enabling mobilisation. *Plasmid* 2017; **89**: 42-8.
48. Manges AR, Mende K, Murray CK et al. Clonal distribution and associated characteristics of Escherichia coli clinical and surveillance isolates from a military medical center. *Diagnostic Microbiology and Infectious Disease* 2017; **87**: 382-5.
49. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)* 2011; **27**: 1009-10.

Chapter 7: General Discussion

In recent years, whole genome sequencing has become integral for molecular analyses, with ease of access increasing and the price of short read sequencing technologies constantly declining. While long read technologies such as PacBio are less readily available and more expensive, they are essential in the production of closed genomes and the investigation of genetic contexts of acquired antimicrobial resistance and virulence determinants. The ability to multiplex up to 24 genomes on long read sequencing devices is having a significant impact on reducing the cost of long read sequencing and it is expected that these technologies will soon become mainstream. Both technologies were used during my candidature; short read Illumina technology for high throughput whole genome sequencing and bioinformatic analyses, while two MDR strains from emerging sequence types housing IS*26*-mediated genetic signatures and complex antimicrobial resistance loci were sequenced using PacBio. Short read sequencing results in scaffolded genomes in part due to the frequent occurrence of repetitive sequences such as insertion elements. These elements are co-located in regions where antimicrobial resistance genes reside and indeed play an important role in the capture and dissemination of these and other genes. Consequently, antimicrobial resistance regions are often difficult to assemble. However, deep read depths afforded by short read sequencing provides an unparalleled opportunity to conduct phylogenetic analyses, subtyping methods including serotyping, phylotyping and sequence typing, more detailed single nucleotide polymorphism (SNP) analyses and gene presence studies.

Through the use of short read sequencing technology this study aimed to fully type a cohort of 83 class 1 integron positive clinical ExPEC strains, characterising them based not only on their antimicrobial resistance and virulence gene cargo but also on unique signatures formed by insertion element IS*26*. Unique insertion events provide a barcode and as such a means of determining transfer of mobile elements between strains and host species, assessing zoonotic potential and gaining insight into the selective pressures that drive the propagation of mobile elements.

7.1 Dominant Lineages in Clinical ExPEC and Trends in Antibiotic Resistance

Multi-locus sequence typing (MLST) is a portable, highly reproducible and reliable method of genotyping bacteria. It has been a gold standard for almost 20 years and it is a typing scheme that

is digital, highly reproducible, and recognised by clinicians, veterinarians and public health epidemiologists. Serotyping has been used for over 100 years as a method to subclassify bacteria and viruses within a species in an epidemiologically meaningful way and as a means to characterise and track important pathotypes. In the traditional sense it requires skilled staff for data interpretation and requires the use of animals (rabbits) to generate sera. With the advent of cost-effective high throughput short-read, whole genome sequencing and subsequent bioinformatic pipelines, MLST and e-serotyping have become mainstream. Numerous "pandemic" sequence types are circulating within ExPEC from clinical settings[43]. As such, a research bias exists towards these sequence types ignoring those that may be emerging. The cohort of genomes sequenced during this candidature, although representing a small snapshot of one hospital environment in Sydney, correlates with current ExPEC population data, with the major sequence types represented being ST95 (17.6%), ST73 (17.6%) and ST131 (12.9%) (See Figure 7.1 for distribution breakdown). Strains were chosen from a much larger collection of ExPEC based on class 1 integron carriage, a reliable proxy for multiple drug resistance, however this did not appear to affect sequence type distribution, an observation consistent with overseas reports noting an increase in MDR status in major *E. coli* STs[136-138]. The strength of this dataset is that we were able to examine the sequences of these human UTI isolates and make insightful comparisons with integron carrying MDR *E. coli* from commensals and pathogenic *E. coli* from food animal production.

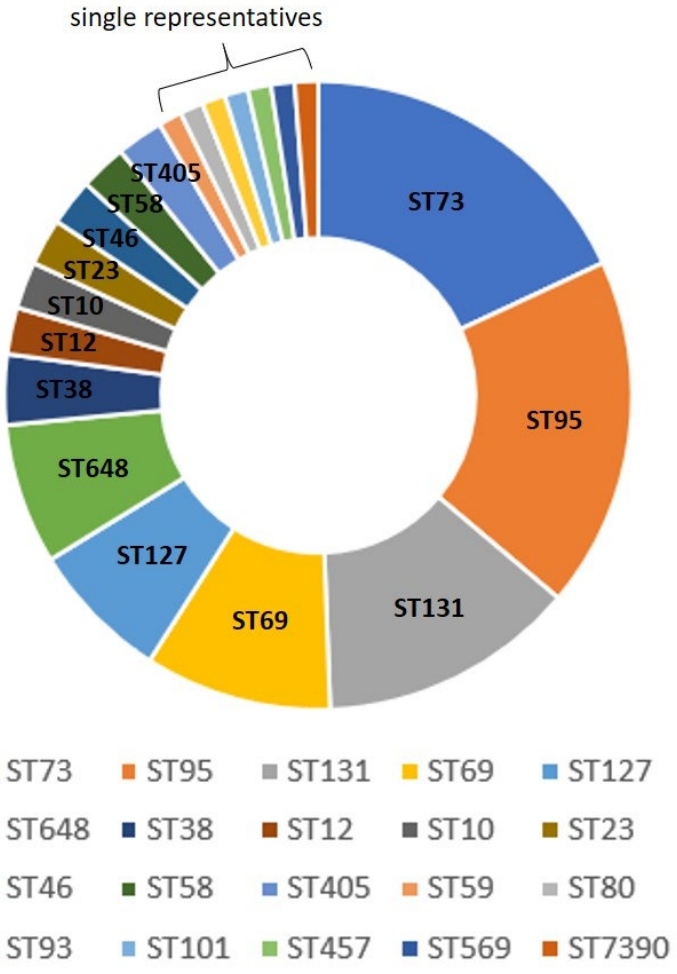


Figure 7.1: Pie chart showing distribution of sequence types amongst cohort. All ST from ST59 onward were singularly represented.

Sequence type composition of the cohort while expected is an important reflection on the current state of *E. coli* causing human disease. The major representative, ST95, while considered "ExPEC" also falls under the sub-pathotype avian pathogenic *E. coli* and causes collibacillosis in domestic and wild birds[126, 139-141] as well as neonatal meningitis[142, 143], blood-stream infection and urinary tract infection[141, 144, 145]. Generally, ST95 shows a low prevalence of antimicrobial resistance as the most dominant *fimH* sublineage, *fimH6*, is frequently pansusceptible[146] suggesting that multidrug resistance is not a defining factor of the success of this clone. However, of the fifteen ST95 strains sequenced during my candidature, only four housed 3 or fewer antibiotic resistance genes and three of these strains were typed as *fimH6* while the more resistant strains mainly grouped into the *fimH1* sublineage.

In terms of antibiotic resistance genes, there was significant variation within most sequence types but while there was no conserved set of genes defining any one ST, certain trends were observed (Table 7.1). Excluding STs which did not have enough representatives to discern trends, the most conserved ST was ST73, with all strains harbouring genes encoding resistance to aminoglycosides, β-lactams and sulphonamides with very few strains containing any genotypic resistance to the other classes screened. ST648 followed closely with all strains containing genes encoding resistance to aminoglycosides, β-lactams, macrolides, sulphonamides and tetracycline but with only half the subset containing genes encoding phenicol and quinolone resistance. Regarding the full cohort, most common genotypic resistances observed were to aminoglycosides, β-lactams and sulphonamides which is likely explained by the presence of class 1 integrons containing *aadA* and *sul1* genes, and possible presence of Tn*6029* whose structure includes $bla_{\text{TEM-1}}$, *sul2* and *strAB* accounting for all three classes on its own. While we were unable to assemble Tn*6029* using short read sequencing all strains containing $bla_{\text{TEM-1}}$, *sul2* and *strAB* are consistent with carriage of these transposons. There was comparably less resistance to last line antibiotics amongst the cohort.

Strains were screened for over 70 putative virulence factors from a curated in-house database spanning adhesins, copper resistance genes, iron acquisition and salmochelin production genes, haemolysins, serum survival proteins, toxins, colicins, invasins, membrane proteases. As with antibiotic resistance no conserved sets of genes were observed within STs, however some STs were considerably more virulent. ST95 and ST73, globally disseminated and established ExPEC STs and the two most highly represented of the cohort, as well as ST127, an emerging ST far less represented in the literature than other lineages, contained the most putative virulence associated genes at an average of 40.2, 39.2 and 38 genes per strain respectively. ST131, a pandemic lineage documented namely for its multidrug resistance phenotype frequently harbouring ESBL genes, averaged 30.5 genes per strain. The most frequently occurring putative virulence factors were iron acquisition and salmochelin production genes, unsurprising as *E. coli* requires iron for metabolic processes in host niches and the majority of the cohort was isolated from urine, a free iron starved environment[147].

Table 7.1: Number of strains with genotypic resistance to antibiotic classes amongst STs. Truncations were considered as absence.

ST	Number of Strains	Sites	Number of strains with genotypic resistance to antibiotic class						
			Aminoglycosides	β-lactams	Macrolides	Phenicols	Quinolones	Sulphonamides	Tetracycline
10	2	wound, urine	2	0	1	1	0	2	1
12	2	urine, blood	2	1	0	0	0	2	0
23	2	urine, blood	0	0	0	0	0	0	0
38	3	urine, blood	3	3	3	0	1	3	1
46	2	urine	2	2	0	1	0	2	1
58	2	urine, blood	2	2	0	0	0	2	2
59	1	urine	1	1	0	1	0	1	0
69	8	urine	8	7	3	0	0	8	3
73	15	urine	15	15	0	1	0	15	2
80	1	urine	1	0	0	0	0	1	0
93	1	urine	1	1	0	1	1	1	1
95	15	urine, blood, wound, vagina	14	14	0	10	0	14	11
101	1	urine	1	0	0	0	0	1	1
127	6	urine, vagina	6	6	1	1	0	6	2
131	11	urine, sputum	11	9	9	2	3	11	4
405	2	urine	2	2	2	0	0	2	0
457	1	urine	1	1	1	1	0	1	0
569	1	urine	1	1	1	1	0	1	1
648	6	urine, blood	6	6	6	3	3	6	6
7390	1	urine	1	1	0	0	0	1	0
Total	83		80	72	27	23	8	80	36

7.2 IS*26*-mediated Insertion Signatures and Zoonosis

Unique and epidemiologically important molecular signatures can be created when an insertion element associated with a resistance gene region induces a deletion event. These unique events provide a means to track the movement of the element that carries multiple drug resistance genes (e.g. a larger transposon, a mobilisable or conjugative plasmid, or integrative conjugative element) between strains, host species and across wide geographical areas, providing insight into the selection pressures that drive their dissemination and establishment. Strains in the collection were interrogated for IS*26*-mediated deletions within clustered resistance gene regions located in the near vicinity of class 1 integrons. These regions were of interest because they provide the host bacterium with two unrelated mechanisms to capture antibiotic resistance gene cargo and enable the building of larger repositories of resistance genes. We have shown how these molecular signatures can be used to track MDR pathogens[11, 62, 99, 100, 113, 127]. As well as studies presented in this thesis, Roy Chowdhury et al. 2015[113] shows this idea in practice, investigating the evolution of a CRL housed by a genomic island within O104:H4 *E. coli* strains isolated from a HUS outbreak in Germany and a patient with bloody diarrhoea in the Republic of Georgia by using a molecular signature formed by IS*26* insertion. These strains shared this signature with another strain isolated from the stool of an apparently healthy person in Nigeria in 2005, resident on a plasmid designated pASL01a[148]. It is clear from information gleaned in this study that this particular signature is globally disseminated and readily trackable, providing insight into selection of mobile elements based on their conferred survival benefits to the host and highlighting the applications of investigating unique molecular signatures.

Long read sequencing is necessary to fully appreciate how resistance genes cluster on mobile genetic elements. Two strains (2009-27 – Chapter 3 and 2009-52 – Chapter 5) housing such signatures underwent PacBio sequencing to fully ascertain complex resistance structures as well as their genetic context.

ST405 strain 2009-27 contained a large chromosomally located IS*26* composite transposon, complete with direct repeats formed upon insertion suggesting a recent transposition to this site. The complex resistance structure, termed Tn*6242*, encodes resistance to five antibiotic classes

and contains five IS*26* elements which have instigated a complicated evolutionary pathway to result in the final structure (see Chapter 3). Tn*6242* contains two IS*26* signature insertions, one causing a truncation of the class 1 integrase gene, and one within the *sul1* gene, not causing a deletion but instead splitting the gene rendering it non-functional. This is the first published instance of this insertion site, however from interrogation of the RefSeq database it is evident that this transposon and its molecular signatures are not confined to Australia, with transposon Tn*6242* present in two other ST405 strains sourced from Danish patients[62](Chapter 3). This analysis exemplifies the significance of utilising molecular signatures for the tracking of antimicrobial resistance structures as Tn*6242*, given its chromosomal positioning and the presence of direct repeats, is both vertically inheritable and horizontally transferrable. The composite transposon and others like it are a reservoir for resistance genes, with the complete structure as well as each IS*26* bounded segment therein being independently mobile and capable of transposition to plasmids, which could consequentially result in transfer to disparate species and hosts. As well as this, each IS*26* element serves as an insertion point for subsequent IS*26* associated translocatable units[118].

The class 1 integron IS*26* insertion point within Tn*6242* while an example of widespread distribution of a signature, has thus far only been observed in strains from human clinical infection. However, the most prevalent IS*26* mediated 3´CS deletion in the cohort is evidence of cross-species transfer between human and animal niches. Paired ST58 strains 2009-49 and 2009-52 (Chapter 5) isolated from urinary tract infection and subsequent urosepsis housed a class 1 integron with an IS*26* insertion 24 bp into the integron 3´CS resulting in the deletion of the *qacEΔ1* and *sul1* genes. This signature deletion appeared in 5 other strains in this cohort including two ST95, and ST12 and two ST23. Regarding strain 2009-52 this IS*26* is the first of three IS*26* elements forming composite transposon Tn*6029*, containing genes conferring resistance to ampicillin, streptomycin and sulphonamides. The entire class 1 integron structure encodes resistance to five antibiotic classes as well as mercury and can transpose as a single element owing to the Tn*1721*/Tn*21* hybrid transposon flanking its limits, evidenced by the 5 bp direct repeats formed by its insertion. This signature insertion has been previously reported in human[11, 61, 99], bovine[99] and porcine sourced strains[100], and as this is a unique evolutionary event it can be surmised that this structure has transferred between host species via lateral gene transfer. With knowledge of this complex resistance loci permeating both human clinical and

veterinary niches, it is crucial to determine the means of transfer through investigation of context. In the case of strain 2009-52, the structure was harboured by a large IncF ColV virulence plasmid[61, 127] as revealed by long-read sequencing. This plasmid, termed pSDJ2009-52F, contained extensive virulence attributes including colicins, iron acquisition and salmochelin production systems, increased serum survival and a haemolysin as well as macrolide-specific efflux proteins. It is clear from this array of pathogenicity determinants that this plasmid likely contributed to the ability of the strain to survive in the urinary tract and subsequently invade the bloodstream causing significant morbidity as ST58 is a commensal phylogroup B1 *E. coli*.

While the class 1 integron containing the IS*26* signature insertion has been reported in multiple host species, the IncF ColV plasmid backbone has similarly been identified in both human clinical and animal contexts (Chapter 6). Much like ST95, IncF ColV plasmid backbones are generally associated with APEC[149]. Outside of this association however, there are numerous reports of this plasmid backbone in human sourced strains both commensal and pathogenic, canine and porcine sourced strains and concerningly, retail meat[127]. There is a pattern emerging with respect to the role of ColV plasmids in human disease, exemplified by a very recent study indicating that *E. coli* from retail meat is a vehicle of human exposure and infection by strains exhibiting presence of ColV plasmids and therefore likely poultry origin[150]. This study investigated *E. coli* from retail meat products and clinical specimens over a 1 year period with a focus on ST131, finding that the ST131-H22 sublineage has become established in poultry populations and appears to frequently house ColV plasmids with 23 of 25 meat sourced ST131-H22 strains containing one.

While short read sequencing cannot by nature reveal complete regions without extensive investigation and PCR linkage strategies, gene presence/absence can be easily determined provided genome scaffolding is below an acceptable threshold. For this reason, it is difficult to determine the prevalence of IncF ColV plasmid backbones in the larger cohort from WGS data alone, however tentative conclusions can be drawn based on virulence gene cargo and plasmid typing. Fourteen strains contained identical virulence gene profiles to pSDJ2009-52F in conjunction with at least one F-type plasmid replicon. The ST distribution alludes to APEC origin or transfer as nine of these belong to ST95, with ST101, ST131, ST46 and ST23 also represented. Interestingly, all nine ST95 strains as well as the single ST131 strain contained

IS*26*-mediated deletions of the macrolide resistance gene *mefB*. While as previously stated it is difficult to determine whether the *mefB* gene is associated with the IncF colV plasmid backbone or even whether the backbone exists in these strains from WGS data, this correlation is nonetheless intriguing and requires further investigation.

The *mefB* gene, generally associated with *sul3* class 1 integrons (Figure 7.2), appears to be a common site for IS*26* insertion events, and different insertion points seem dependent on host species. A *mefB* gene fragment of 890 bp had the highest instance in this cohort and occurred only in ST95 strains; nine contained all genes necessary for potential carriage of an IncF colV plasmid, and two contained all but one. This fragment has not been specifically reported, however an interrogation of GenBank found that there is only one instance of this deletion occurring outside of this cohort, on an IncF colV plasmid in a commensal ST95 strain from an apparently healthy human sampled in Sydney, Australia in 2010 [85]. It is clear from this information that thus far the 890 bp *mefB* fragment is human sourced *E. coli* specific, though not limited to pathogenic strains, however this could be the result of bias towards the isolation and characterisation of strains of *E. coli* carrying clinically important antimicrobial resistance genes such as those expressing ESBLs and the dismissal of truncated non-functional genes as relevant. Two other *mefB* signature deletions were represented in the cohort; a 714 bp fragment entirely unique to this collection, and a 111 bp fragment which has been observed in a porcine collection spanning 13 STs[100]. Both were singularly represented, with the 111 bp fragment present in a strain belonging to globally established ST131, well documented for being predominantly responsible for global dissemination of fluoroquinolone and cephalosporin resistance[151-153].

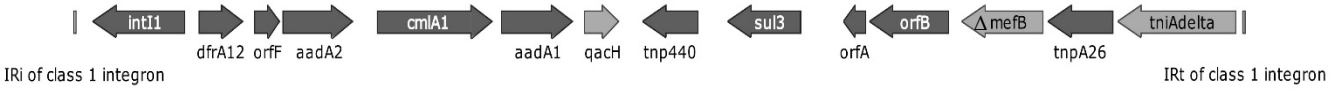


Figure 7.2: Structure of a *sul3* integron, associated with the *mefB* gene, which has been truncated to 890 bp by IS*26*.

7.3 Concluding Statements

The observation of a significant number of CRL formed by IS*26* mediated evolutionary events in this cohort emphasizes the potential of molecular signatures to be used in the monitoring and surveillance of the lateral gene transfer of antibiotic resistance regions. Signatures and mobile elements described here are strong evidence for transfer of genetic material between bacterial populations (Chapter 6), host species (Chapter 5), disparate geographical areas (Chapters 3, 5 and 6) and *E. coli* lineages (above).

It is clear from this study and the literature that selective pressures applied in human medicine, veterinary practices and agriculture as well as the environment have a serious impact on the formation and dissemination of complex resistance structures and plasmids.

In Australia, sound antimicrobial stewardship practices have shown that carriage of antibiotic genes encoding resistance to clinically important antibiotics are infrequently detected[100, 126, 154]. However, on farms where ceftiofur has been used to control problematic infections or used off label, there is a clear precedence for selection of CTX-M-1 and other genes encoding extended spectrum B-lactams. One study showed that carriage of CTX-M-1 persisted for more than 4 years after cessation of use on a farm[154]. A common argument against this is that antibiotics used in a human clinical context are not generally used in veterinary practices and vice versa. Due to restrictions and regulations on antibiotic stewardship in many cases this is true, however this argument ignores co-selection of antimicrobial resistance and the mosaic structures that arise from selection pressures (Figure 7.3).

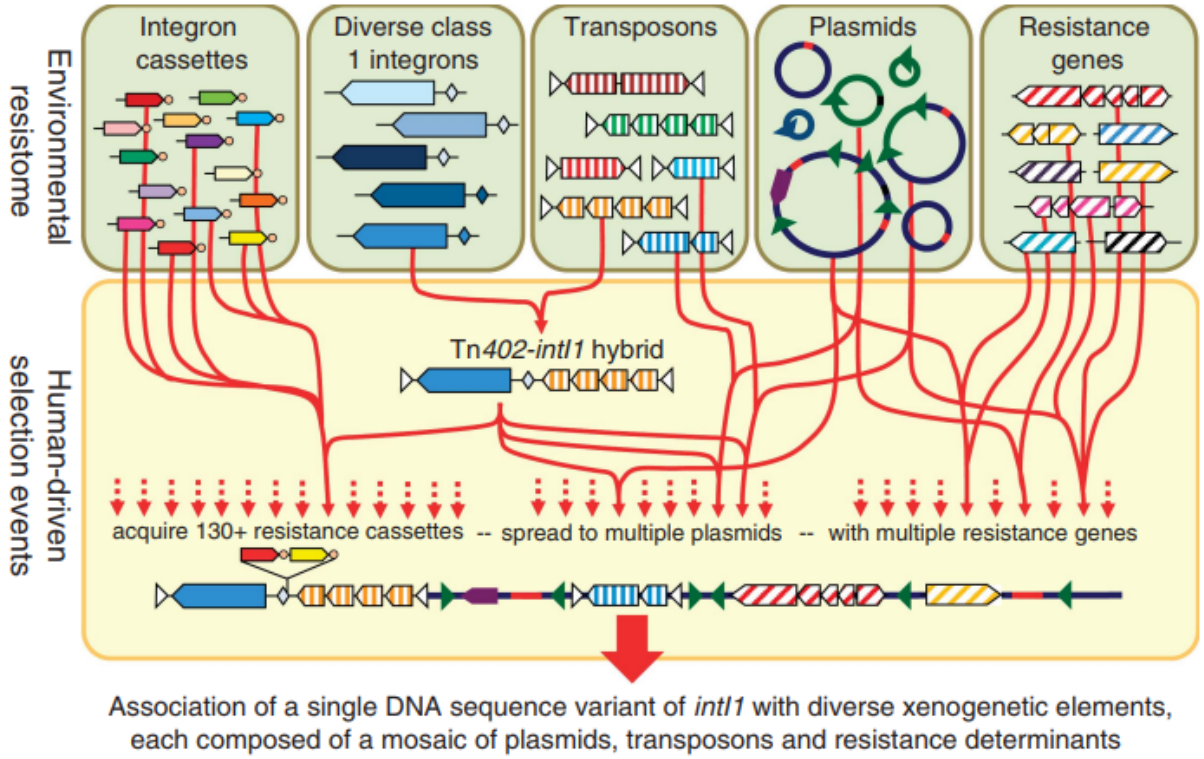


Figure 7.3: Diagrammatical representation from Gillings *et al*. 2015 of the ways in which the environmental mobilome assembles complex mosaic antimicrobial resistance structures in response to selection pressures[124].

Tn*6026*, an amalgamation of transposon Tn*6029* mentioned previously and Tn*4352*, carries both the ampicillin resistance gene $bla_{\text{TEM-1}}$ which is relevant to human medicine, as well as the *aphA1* gene encoding resistance to neomycin and kanamycin, an antibiotic used in treatment and prophylaxis of livestock, particularly swine. Additionally, this transposon has been reported in association with class 1 integrons housing the *dfrA5* gene which encodes resistance to trimethoprim, used empirically on a global scale to treat urinary tract infection, as well as mercury resistance transposons, significant as mercury is ubiquitous in environmental settings[75, 99, 125, 155]. The correct selection pressure applied in any one niche (Figure 1.1) would drive the propagation of this structure, having implications for the remaining niches. Selection pressures are also amplified due to crossover between environments, namely anthropogenic pollution; heavy metals, pharmaceuticals and domestic, industrial and agricultural waste produced by

humans entering the environment. For example, between 30 and 90% of human and animal ingested antibiotics are excreted unmodified[156], and up to 80% of antibiotics used in aquaculture reach the environment[157]. This results in a devastating amount of active antibiotic compounds reaching the natural environment via human activities and waste production.

Studies such as those conducted during my candidature involving large sample pools and in-depth analyses provide a greater understanding of the mechanisms and driving forces behind antimicrobial resistance, as well as deliver a basis for implementation of strategies to reduce the impact and spread of antibiotic resistance.

7.3.1 Limitations

Eighty-three class 1 integrons containing clinical strains underwent whole genome sequencing during this project, with the intention of detailed phylogenetic analyses and investigation of antimicrobial and virulence gene cargo and the elements that mobilise it. While short read sequencing is appropriate for phylogenetic analysis and gene screening, limitations exist in determining the genetic context of antimicrobial resistance genes and virulence factors. This was achieved with strains 2009-27 and 2009-52 through use of long read sequencing technologies, which resulted in closed genomes and circularised plasmids. While crucial in ascertaining genetic context, closed genomes resulting from long read sequencing can also be used as reference genomes for comparative analyses such as phylogenetics and single nucleotide polymorphisms. This technology is currently too expensive to be applied to large collections, however it is still very much under development.

This collection is also representative of only a single Sydney hospital. Findings could be validated through expanded sampling; multiple hospitals over a wide geographical area, for a fuller understanding of the landscape of antimicrobial resistance and virulence within clinical *E. coli* in Australia. Moreover, interrogation of class 1 integron negative strains may expose ancestral strains or plasmids, which would provide valuable insight into the microevolutionary events leading to the complex resistance loci described in this thesis.

Conversely, studies focussing on antibiotic resistance in UPEC populations generally ignore the associated epidemiological data i.e. phylogenetic analyses and sequence typing. An expanded and well-rounded approach to the study of UPEC populations could provide invaluable insights into the emergence or establishment of strains and their related antibiotic resistance, contributing positively to efforts in preventing further damage caused by indiscreet antibiotic usage.

References

1. D'Costa VM, King CE, Kalan L et al. Antibiotic resistance is ancient. *Nature* 2011; **477**: 457-61.
2. Schlacher K, Goodman MF. Lessons from 50 years of SOS DNA-damage-induced mutagenesis. *Nature reviews Molecular cell biology* 2007; **8**: 587-94.
3. Tenaillon O, Denamur E, Matic I. Evolutionary significance of stress-induced mutagenesis in bacteria. *Trends in microbiology* 2004; **12**: 264-70.
4. Berkelman RL. Consideration of prophylactic antibiotic therapy during an outbreak of Legionnaires' disease. *Clin Infect Dis* 2014; **58**: 144-5.
5. Klugman KP, Izadnegahdar R. Antibiotic prophylaxis-Preventing severe infections and saving lives in poor countries with very high mortality risk. *PLoS medicine* 2018; **15**: e1002594.
6. Chan YH, Fan MM, Fok CM et al. Antibiotics nonadherence and knowledge in a community with the world's leading prevalence of antibiotics resistance: implications for public health intervention. *Am J Infect Control* 2012; **40**: 113-7.
7. Doyle MP, Schoeni JL. Isolation of Escherichia coli O157:H7 from retail fresh meats and poultry. *Appl Environ Microbiol* 1987; **53**: 2394-6.
8. Rangel JM, Sparling PH, Crowe C et al. Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. *Emerg Infect Dis* 2005; **11**: 603-9.
9. Tilden J, Jr., Young W, McNamara AM et al. A new route of transmission for Escherichia coli: infection from dry fermented salami. *American journal of public health* 1996; **86**: 1142-5.
10. Zhao S, White DG, McDermott PF et al. Identification and expression of cephamycinase bla(CMY) genes in Escherichia coli and Salmonella isolates from food animals and ground meat. *Antimicrob Agents Chemother* 2001; **45**: 3647-50.
11. 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.
12. Roy Chowdhury P, McKinnon J, Wyrsch E et al. Genomic interplay in bacterial communities: implications for growth promoting practices in animal husbandry. *Front Microbiol* 2014; **5**: 394.
13. Doyle ME, Archer J, Kaspar CW et al. Human Illness Caused by E. coli O157:H7 from Food and Non-food Sources.
14. Alanis AJ. Resistance to antibiotics: are we in the post-antibiotic era? *Archives of medical research* 2005; **36**: 697-705.
15. Morales E, Cots F, Sala M et al. Hospital costs of nosocomial multi-drug resistant Pseudomonas aeruginosa acquisition. *BMC health services research* 2012; **12**: 122.
16. Markogiannakis H, Pachylaki N, Samara E et al. Infections in a surgical intensive care unit of a university hospital in Greece. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2009; **13**: 145-53.
17. Suwantarat N, Supple LA, Cadnum JL et al. Quantitative assessment of interactions between hospitalized patients and portable medical equipment and other fomites. *Am J Infect Control* 2017; **45**: 1276-8.

18. Di Lodovico S, Cataldi V, Di Campli E et al. Enterococcus hirae biofilm formation on hospital material surfaces and effect of new biocides. *Environmental health and preventive medicine* 2017; **22**: 63.
19. Krzysciak P, Chmielarczyk A, Pobiega M et al. Acinetobacter baumannii isolated from hospital-acquired infection: biofilm production and drug susceptibility. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* 2017; **125**: 1017-26.
20. Allen HK, Donato J, Wang HH et al. Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol* 2010; **8**: 251-9.
21. Bhullar K, Waglechner N, Pawlowski A et al. Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One* 2012; **7**: e34953.
22. Humeniuk C, Arlet G, Gautier V et al. Beta-lactamases of Kluyvera ascorbata, probable progenitors of some plasmid-encoded CTX-M types. *Antimicrob Agents Chemother* 2002; **46**: 3045-9.
23. Boto L. Horizontal gene transfer in evolution: facts and challenges. *Proceedings Biological sciences* 2010; **277**: 819-27.
24. Furuya EY, Lowy FD. Antimicrobial-resistant bacteria in the community setting. *Nat Rev Microbiol* 2006; **4**: 36-45.
25. Zhaxybayeva O, Doolittle WF. Lateral gene transfer. *Current biology : CB* 2011; **21**: R242-6.
26. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000; **405**: 299-304.
27. Popa O, Dagan T. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* 2011; **14**: 615-23.
28. Kenzaka T, Nasu M, Tani K. Transfer of a phage T4 gene into Enterobacteriaceae, determined at the single-cell level. *Appl Environ Microbiol* 2010; **76**: 1274-7.
29. Halary S, Leigh JW, Cheaib B et al. Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci U S A* 2010; **107**: 127-32.
30. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 2005; **3**: 711-21.
31. Alberts B, Johnson A, Lewis J et al. *Molecular Biology of the Cell*. New York: Garland Science, 2002.
32. Mandali S, Dhar G, Avliyakulov NK et al. The Site-Specific Integration Reaction of Listeria Phage A118 Integrase, a Serine Recombinase. *Mobile DNA* 2013; **4**.
33. Smith MCA, Till R, Brady K et al. Synapsis and DNA Cleavage in phiC31 Integrase-Mediated Site-Specific Recombination. *Nucleic acids research* 2004; **32**: 2607-17.
34. Unger F, Eisenberg T, Prenger-Berninghoff E et al. Imported reptiles as a risk factor for the global distribution of Escherichia coli harbouring the colistin resistance gene mcr-1. *Int J Antimicrob Agents* 2017; **49**: 122-3.
35. Rouli L, Merhej V, Fournier PE et al. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New microbes and new infections* 2015; **7**: 72-85.
36. Pedersen TL. Hierarchical sets: analyzing pangenome structure through scalable set visualizations. *Bioinformatics (Oxford, England)* 2017; **33**: 1604-12.
37. Rasko DA, Rosovitz MJ, Myers GS et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. *J Bacteriol* 2008; **190**: 6881-93.

38. Pang TY, Lercher MJ. Each of 3,323 metabolic innovations in the evolution of E. coli arose through the horizontal transfer of a single DNA segment. *Proc Natl Acad Sci U S A* 2019; **116**: 187-92.
39. Gomes TA, Elias WP, Scaletsky IC et al. Diarrheagenic Escherichia coli. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]* 2016; **47 Suppl 1**: 3-30.
40. Weichhart T, Haidinger M, Hörl WH et al. Current concepts of molecular defence mechanisms operative during urinary tract infection. *European Journal of Clinical Investigation* 2008; **38**: 29-38.
41. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *The American journal of medicine* 2002; **113 Suppl 1A**: 5s-13s.
42. Manges AR, Tabor H, Tellis P et al. Endemic and epidemic lineages of Escherichia coli that cause urinary tract infections. *Emerg Infect Dis* 2008; **14**: 1575-83.
43. Riley LW. Pandemic lineages of extraintestinal pathogenic Escherichia coli. *Clin Microbiol Infect* 2014; **20**: 380-90.
44. Croxall G, Hale J, Weston V 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**: 2501-8.
45. Tartof SY, Solberg OD, Manges AR et al. Analysis of a uropathogenic Escherichia coli clonal group by multilocus sequence typing. *J Clin Microbiol* 2005; **43**: 5860-4.
46. Campos ACC, Andrade NL, Ferdous M et al. Comprehensive Molecular Characterization of Escherichia coli Isolates from Urine Samples of Hospitalized Patients in Rio de Janeiro, Brazil. *Front Microbiol* 2018; **9**: 243.
47. Yamaji R, Rubin J, Thys E et al. Persistent Pandemic Lineages of Uropathogenic Escherichia coli in a College Community from 1999 to 2017. *J Clin Microbiol* 2018; **56**.
48. Matsumura Y, Yamamoto M, Nagao M et al. Emergence and spread of B2-ST131-O25b, B2-ST131-O16 and D-ST405 clonal groups among extended-spectrum-beta-lactamase-producing Escherichia coli in Japan. *J Antimicrob Chemother* 2012; **67**: 2612-20.
49. Matsumura Y, Yamamoto M, Nagao M et al. Association of fluoroquinolone resistance, virulence genes, and IncF plasmids with extended-spectrum-beta-lactamase-producing Escherichia coli sequence type 131 (ST131) and ST405 clonal groups. *Antimicrob Agents Chemother* 2013; **57**: 4736-42.
50. Kudinha T, Johnson JR, Andrew SD et al. Escherichia coli sequence type 131 as a prominent cause of antibiotic resistance among urinary Escherichia coli isolates from reproductive-age women. *J Clin Microbiol* 2013; **51**: 3270-6.
51. Petty NK, Ben Zakour NL, Stanton-Cook M et al. Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A* 2014; **111**: 5694-9.
52. Price LB, Johnson JR, Aziz M 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**: e00377-13.
53. Totsika M, Beatson SA, Sarkar S 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**: e26578.

54. Christiansen N, Nielsen L, Jakobsen L et al. Fluoroquinolone resistance mechanisms in urinary tract pathogenic Escherichia coli isolated during rapidly increasing fluoroquinolone consumption in a low-use country. *Microb Drug Resist* 2011; **17**: 395-406.
55. Cullen IM, Manecksha RP, McCullagh E et al. An 11-year analysis of the prevalent uropathogens and the changing pattern of Escherichia coli antibiotic resistance in 38,530 community urinary tract infections, Dublin 1999-2009. *Irish journal of medical science* 2013; **182**: 81-9.
56. Johnson JR, Johnston B, Clabots C et al. Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant E. coli infections in the United States. *Clin Infect Dis* 2010; **51**: 286-94.
57. Coque TM, Novais A, Carattoli A et al. Dissemination of clonally related Escherichia coli strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerg Infect Dis* 2008; **14**: 195-200.
58. Nicolas-Chanoine MH, Blanco J, Leflon-Guibout V et al. Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* 2008; **61**: 273-81.
59. Rogers BA, Sidjabat HE, Paterson DL. Escherichia coli O25b-ST131: a pandemic, multiresistant, community-associated strain. *J Antimicrob Chemother* 2011; **66**: 1-14.
60. Bogema DR, McKinnon J, Liu M 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. *Microbial genomics* 2019.
61. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. *Int J Antimicrob Agents* 2018; **52**: 430-5.
62. Roy Chowdhury P, McKinnon J, Liu M et al. Multidrug Resistant Uropathogenic Escherichia coli ST405 With a Novel, Composite IS26 Transposon in a Unique Chromosomal Location. *Frontiers in Microbiology* 2019; **9**.
63. Delannoy S, Beutin L, Fach P. Discrimination of enterohemorrhagic Escherichia coli (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. *J Clin Microbiol* 2013; **51**: 3257-62.
64. Mead PS, Slutsker L, Dietz V et al. Food-related illness and death in the United States. *Emerg Infect Dis* 1999; **5**: 607-25.
65. Novicki TJ, Daly JA, Mottice SL et al. Comparison of Sorbitol MacConkey Agar and a Two-Step Method Which Utilizes Enzyme-Linked Immunosorbent Assay Toxin Testing and a Chromogenic Agar To Detect and Isolate EnterohemorrhagicEscherichia coli. *Journal of Clinical Microbiology* 2000; **38**: 547-51.
66. Bielaszewska M, Mellmann A, Bletz S et al. Enterohemorrhagic Escherichia coli O26:H11/H-: a new virulent clone emerges in Europe. *Clin Infect Dis* 2013; **56**: 1373-81.
67. Willey JM, Sherwood LM, Woolverton CJ. *Prescott, Harley and Klein's Microbiology*. New York: McGraw-Hill, 2008.
68. Carter AO, Borczyk AA, Carlson JA et al. A severe outbreak of Escherichia coli O157:H7--associated hemorrhagic colitis in a nursing home. *N Engl J Med* 1987; **317**: 1496-500.
69. Heiman KE, Mody RK, Johnson SD et al. Escherichia coli O157 Outbreaks in the United States, 2003-2012. *Emerg Infect Dis* 2015; **21**: 1293-301.
70. Jenkins C, Dallman TJ, Grant KA. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing Escherichia coli serogroup

O157:H7, England, 2013 to 2017. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2019; **24**.
71. Riley LW, Remis RS, Helgerson SD et al. Hemorrhagic colitis associated with a rare Escherichia coli serotype. *N Engl J Med* 1983; **308**: 681-5.
72. Wells JG, Davis BR, Wachsmuth IK et al. Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare Escherichia coli serotype. *J Clin Microbiol* 1983; **18**: 512-20.
73. Frenzen PD, Drake A, Angulo FJ. Economic cost of illness due to Escherichia coli O157 infections in the United States. *Journal of food protection* 2005; **68**: 2623-30.
74. Copur-Cicek A, Ozgumus OB, Saral A et al. Antimicrobial resistance patterns and integron carriage of Escherichia coli isolates causing community-acquired infections in Turkey. *Annals of laboratory medicine* 2014; **34**: 139-44.
75. Cain AK, Liu X, Djordjevic SP et al. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant Salmonella enterica serovar Typhimurium from Australian animals. *Microb Drug Resist* 2010; **16**: 197-202.
76. Harmer CJ, Moran RA, Hall RM. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* 2014; **5**: e01801-14.
77. Leplae R, Hebrant A, Wodak SJ et al. ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic acids research* 2004; **32**: D45-9.
78. Frost LS, Leplae R, Summers AO et al. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005; **3**: 722-32.
79. Heinemann JA, Sprague GF, Jr. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* 1989; **340**: 205-9.
80. Partridge SR, Kwong SM, Firth N et al. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev* 2018; **31**.
81. Silva F, Queiroz JA, Domingues FC. Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnology advances* 2012; **30**: 691-708.
82. Sengupta M, Austin S. Prevalence and Significance of Plasmid Maintenance Functions in the Virulence Plasmids of Pathogenic Bacteria. *Infection and Immunity* 2011; **79**: 2502-9.
83. San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiol Spectr* 2017; **5**.
84. Carattoli A, Zankari E, Garcia-Fernandez A et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; **58**: 3895-903.
85. 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; **84-85**: 11-9.
86. Hayes F. Transposon-based strategies for microbial functional genomics and proteomics. *Annual review of genetics* 2003; **37**: 3-29.
87. Partridge SR. Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiol Rev* 2011; **35**: 820-55.
88. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British journal of pharmacology* 2008; **153 Suppl 1**: S347-57.
89. Ke X, Gu B, Pan S et al. Epidemiology and molecular mechanism of integron-mediated antibiotic resistance in Shigella. *Archives of microbiology* 2011; **193**: 767-74.

90. Gillings MR, Holley MP, Stokes HW. Evidence for dynamic exchange of qac gene cassettes between class 1 integrons and other integrons in freshwater biofilms. *FEMS Microbiol Lett* 2009; **296**: 282-8.
91. Gillings MR, Holley MP, Stokes HW et al. Integrons in Xanthomonas: a source of species genome diversity. *Proc Natl Acad Sci U S A* 2005; **102**: 4419-24.
92. Michael CA, Gillings MR, Holmes AJ et al. Mobile gene cassettes: a fundamental resource for bacterial evolution. *The American naturalist* 2004; **164**: 1-12.
93. Stokes HW, Holmes AJ, Nield BS et al. Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl Environ Microbiol* 2001; **67**: 5240-6.
94. Boucher Y, Labbate M, Koenig JE et al. Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends in microbiology* 2007; **15**: 301-9.
95. Cambray G, Guerout AM, Mazel D. Integrons. *Annual review of genetics* 2010; **44**: 141-66.
96. Rao AN, Barlow M, Clark LA et al. Class 1 integrons in resistant Escherichia coli and Klebsiella spp., US hospitals. *Emerg Infect Dis* 2006; **12**: 1011-4.
97. Nandi S, Maurer JJ, Hofacre C et al. Gram-positive bacteria are a major reservoir of Class 1 antibiotic resistance integrons in poultry litter. *Proc Natl Acad Sci U S A* 2004; **101**: 7118-22.
98. Domingues S, da Silva GJ, Nielsen KM. Integrons: Vehicles and pathways for horizontal dissemination in bacteria. *Mobile genetic elements* 2012; **2**: 211-23.
99. Dawes FE, Kuzevski A, Bettelheim KA 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**: e12754.
100. Reid CJ, Wyrsch ER, Roy Chowdhury P et al. Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microbial genomics* 2017; **3**.
101. Stokes HW, Hall RM. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* 1989; **3**: 1669-83.
102. Alberts B. DNA replication and recombination. *Nature* 2003; **421**: 431-5.
103. Wirth T, Falush D, Lan R et al. Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol Microbiol* 2006; **60**: 1136-51.
104. Stokes HW, O'Gorman DB, Recchia GD et al. Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* 1997; **26**: 731-45.
105. Michael CA, Labbate M. Gene cassette transcription in a large integron-associated array. *BMC genetics* 2010; **11**: 82.
106. Labbate M, Roy Chowdhury P, Stokes HW. A class 1 integron present in a human commensal has a hybrid transposition module compared to Tn402: evidence of interaction with mobile DNA from natural environments. *Journal of bacteriology* 2008; **190**: 5318-27.
107. Betteridge T, Partridge SR, Iredell JR et al. Genetic context and structural diversity of class 1 integrons from human commensal bacteria in a hospital intensive care unit. *Antimicrob Agents Chemother* 2011; **55**: 3939-43.
108. Kholodii GY, Mindlin SZ, Bass IA et al. Four genes, two ends, and a res region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a mer operon or an integron. *Mol Microbiol* 1995; **17**: 1189-200.
109. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. *Microbiology and molecular biology reviews : MMBR* 1999; **63**: 507-22.

110. Juhas M, van der Meer JR, Gaillard M et al. Genomic Islands: Tools of Bacterial Horizontal Gene Transfer and Evolution. *FEMS Microbiol Rev* 2008; **33**: 376-93.
111. Hacker J, Carniel E. Ecological Fitness, Genomic Islands and Bacterial Pathogenicity. A Darwinian View of the Evolution of Microbes. *EMBO Rep* 2001; **2**: 376-81.
112. Boyd DA, Peters GA, Ng L et al. Partial characterization of a genomic island associated with the multidrug resistance region of Salmonella enterica Typhymurium DT104. *FEMS Microbiol Lett* 2000; **189**: 285-91.
113. 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**: e0115781.
114. McKinnon J, Roy Chowdhury P, Djordjevic SP. Whole genome sequencing of 83 class 1 integron positive clinical isolates from urinary tract infection. Manuscript in preparation.
115. Venturini C, Hassan KA, Roy Chowdhury P 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**: e78862.
116. Cain AK, Hall RM. Evolution of IncHI2 plasmids via acquisition of transposons carrying antibiotic resistance determinants. *J Antimicrob Chemother* 2012; **67**: 1121-7.
117. He S, Hickman AB, Varani AM et al. Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition. *MBio* 2015; **6**: e00762.
118. Harmer CJ, Hall RM. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* 2016; **1**.
119. Melvold JA, Wyrsch ER, McKinnon J et al. Identification of a novel qnrA allele, qnrA8, in environmental Shewanella algae. *J Antimicrob Chemother* 2017; **72**: 2949-52.
120. Toleman MA, Walsh TR. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol Rev* 2011; **35**: 912-35.
121. Poole TL, Callaway TR, Bischoff KM et al. Macrolide inactivation gene cluster mphA-mrx-mphR adjacent to a class 1 integron in Aeromonas hydrophila isolated from a diarrhoeic pig in Oklahoma. *J Antimicrob Chemother* 2006; **57**: 31-8.
122. Gillings MR. Integrons: past, present, and future. *Microbiology and molecular biology reviews : MMBR* 2014; **78**: 257-77.
123. Gillings M, Boucher Y, Labbate M et al. The evolution of class 1 integrons and the rise of antibiotic resistance. *J Bacteriol* 2008; **190**: 5095-100.
124. Gillings MR, Gaze WH, Pruden A et al. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* 2015; **9**: 1269-79.
125. Venturini C, Beatson SA, Djordjevic SP et al. Multiple Antibiotic Resistance Gene Recruitment onto the Enterohemorrhagic Escherichia coli Virulence Plasmid. *FASEB J* 2010; **24**: 1160-6.
126. Cummins ML, Reid C, Roy Chowdhury P et al. Whole genome sequence analysis of Australian avian pathogenic E. coli that carry the class 1 integrase gene. *MGen* 2019.
127. McKinnon J, Roy Chowdhury P, Djordjevic SP. Molecular analysis of an IncF ColV-like plasmid lineage that carries a complex resistance locus with a trackable genetic signature. *Submitted* 2019.
128. Tritt A, Eisen JA, Facciotti MT et al. An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* 2012; **7**: e42304.

129. Hunt M, Silva ND, Otto TD et al. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015; **16**: 294.
130. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. *Journal of molecular biology* 1990; **215**: 403-10.
131. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010; **5**: e11147.
132. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC bioinformatics* 2010; **11**: 595.
133. Darling AE, Jospin G, Lowe E et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014; **2**: e243.
134. Alikhan NF, Petty NK, Ben Zakour NL et al. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011; **12**: 402.
135. Milne I, Stephen G, Bayer M et al. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in bioinformatics* 2013; **14**: 193-202.
136. Alhashash F, Wang X, Paszkiewicz K et al. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR Escherichia coli ST73: high levels of genomic and plasmid diversity in causative isolates. *J Antimicrob Chemother* 2016; **71**: 339-43.
137. Allocati N, Masulli M, Alexeyev MF et al. Escherichia coli in Europe: an overview. *International journal of environmental research and public health* 2013; **10**: 6235-54.
138. Sidjabat HE, Paterson DL. Multidrug-resistant Escherichia coli in Asia: epidemiology and management. *Expert review of anti-infective therapy* 2015; **13**: 575-91.
139. Blanco JE, Blanco M, Mora A et al. Serotypes of Escherichia coli isolated from septicaemic chickens in Galicia (northwest Spain). *Vet Microbiol* 1998; **61**: 229-35.
140. Mora A, Lopez C, Dabhi G et al. Extraintestinal pathogenic Escherichia coli O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. *BMC Microbiol* 2009; **9**: 132.
141. Mora A, Viso S, Lopez C et al. Poultry as reservoir for extraintestinal pathogenic Escherichia coli O45:K1:H7-B2-ST95 in humans. *Vet Microbiol* 2013; **167**: 506-12.
142. Glode MP, Sutton A, Robbins JB et al. Neonatal meningitis due of Escherichia coli K1. *J Infect Dis* 1977; **136 Suppl**: S93-7.
143. Tivendale KA, Logue CM, Kariyawasam S et al. Avian-pathogenic Escherichia coli strains are similar to neonatal meningitis E. coli strains and are able to cause meningitis in the rat model of human disease. *Infect Immun* 2010; **78**: 3412-9.
144. Adams-Sapper S, Diep BA, Perdreau-Remington F et al. Clonal composition and community clustering of drug-susceptible and -resistant Escherichia coli isolates from bloodstream infections. *Antimicrob Agents Chemother* 2013; **57**: 490-7.
145. Stephens CM, Skerker JM, Sekhon MS et al. Complete Genome Sequences of Four Escherichia coli ST95 Isolates from Bloodstream Infections. *Genome Announc* 2015; **3**.
146. Stephens CM, Adams-Sapper S, Sekhon M et al. Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. *mSphere* 2017; **2**.
147. Hancock V, Ferrieres L, Klemm P. The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious Escherichia coli in human urine. *Microbiology (Reading, England)* 2008; **154**: 167-75.
148. Labar AS, Millman JS, Ruebush E et al. Regional dissemination of a trimethoprim-resistance gene cassette via a successful transposable element. *PLoS One* 2012; **7**: e38142.

149. de Oliveira AL, Rocha DA, Finkler F et al. Prevalence of ColV Plasmid-Linked Genes and In Vivo Pathogenicity of Avian Strains of Escherichia coli. *Foodborne Pathog Dis* 2015; **12**: 679-85.
150. Liu CM, Stegger M, Aziz M et al. Escherichia coli ST131-H22 as a Foodborne Uropathogen. *MBio* 2018; **9**.
151. Johnson JR, Tchesnokova V, Johnston B et al. Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. *J Infect Dis* 2013; **207**: 919-28.
152. Nicolas-Chanoine MH, Bertrand X, Madec JY. Escherichia coli ST131, an intriguing clonal group. *Clin Microbiol Rev* 2014; **27**: 543-74.
153. Pitout JD, DeVinney R. Escherichia coli ST131: a multidrug-resistant clone primed for global domination. *F1000Research* 2017; **6**.
154. Abraham S, Jordan D, Wong HS et al. First detection of extended-spectrum cephalosporin- and fluoroquinolone-resistant Escherichia coli in Australian food-producing animals. *Journal of global antimicrobial resistance* 2015; **3**: 273-7.
155. Reid C, Roy Chowdhury P, Djordjevic SP. *Tn6026 and Tn6029 are found in complex resistance regions mobilized by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae*, 2015.
156. Sarmah AK, Meyer MT, Boxall AB. A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. *Chemosphere* 2006; **65**: 725-59.
157. Cabello FC, Godfrey HP, Tomova A et al. Antimicrobial use in aquaculture re-examined: its relevance to antimicrobial resistance and to animal and human health. *Environmental microbiology* 2013; **15**: 1917-42.