Genomic Epidemiology of Multidrug Resistant Porcine Commensal *Escherichia coli*

Cameron James Reid

Bachelor of Biotechnology (Hons)

University of Technology Sydney

Faculty of Science

School of Life Sciences

The i3 Institute

April 2019

Certificate of Authorship

I, Cameron James Reid, declare that this thesis, submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science, School of Life Sciences at the University of Technology Sydney, is wholly my own work unless otherwise 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.

Signature:

Production Note:
Signature removed prior to publication.

Date: 8/4/2019

Acknowledgements

Firstly, I would like to thank my supervisors Professor Steven Djordjevic, Dr Piklu Roy Chowdhury and Dr Toni Chapman for their expertise, training, mentorship and support over the course of my candidature. I consider myself privileged to have had your friendship and guidance over the past 4 years. Additional thanks to AusGem, the NSW Department of Primary Industries and the Elizabeth MacArthur Agricultural Institute for providing me with the excellent collection of *E. coli* I was lucky enough to work with.

I would also like to thank all the members of the Djordjevic lab past and present. In particular I would like to thank Ethan Wyrsch and Jess McKinnon for showing me the ropes and training me in the early stages of my Honours and PhD project.

Thank you to Max Cummins, Tiziana Zingali, Priyanka Hastak Shah and Veronica Jarocki. You have been excellent colleagues and friends. You have made the course of my candidature all the more enjoyable and I will always view this time in my life fondly.

Thank you to the academic and technical staff at the i3 Institute, including Director Professor Liz Harry for the supportive and encouraging environment you have created for young scientists. I hope to see the Institute continue to nurture outstanding researchers and make a difference in the world of infectious disease.

I'd like to thank the University of Technology Sydney for supporting me financially with a Doctoral Scholarship as well as the Australian Government Research Training Program.

Finally, and most importantly, I would like to thank my family and friends. My Mum and Dad in particular have been nothing short of amazing in supporting me emotionally and financially throughout this endeavour. I am truly blessed to have such supportive and encouraging parents who have always believed in me. My brother Max has been a rock, a voice of reason and a shoulder to lean on throughout. I love you all.

Statement

This thesis is by compilation. The first two results chapters are published papers listed in the section below. The final chapter is a publication style manuscript that is being prepared for journal submission. Figure and Table numbers have been edited from the original publications to match the chapter numbering of the thesis. Most of the figures are very large and hard to view as they lose resolution due to the margins required within the thesis. Please view the figures as PDFs attached or in the original publications for Chapters 4 and 5.

List of publications

Paper 1; Chapter 4

Porcine commensal *Escherichia coli*: a reservoir for class 1 integrons associated with IS*26.*

Cameron J. Reid[1†], Ethan R. Wyrsch[1†], Piklu Roy Chowdhury[1], Tiziana Zingali[1], Michael Liu[1], Aaron E. Darling[1], Toni A. Chapman[2], and Steven P. Djordjevic[1]

Microb Genom. 2017 Dec; 3(12): e000143.doi:10.1099/mgen.0.000143

[†]These authors contributed equally to this work.

[1]The i3 Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia

[2]NSW Department of Primary Industries, Elizabeth MacArthur Agricultural Institute, Menangle, NSW 2568, Australia

Paper 2; Chapter 5

Australian porcine clonal complex 10 (CC10) *Escherichia coli* belong to multiple sublineages of a highly diverse global CC10 phylogeny.

Cameron J. Reid[1], Matthew Z. DeMaere[1], Steven P. Djordjevic[1]

Microb Genom. 2018 Oct 10. doi: 10.1099/mgen.0.000225

[1]The i3 Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia

Table of Contents

Abbreviations

AMR	Antimicrobial Resistance
ARG	Antimicrobial Resistance Gene
VAG	Virulence-associated gene
MGE	Mobile Genetic Element
IS	Insertion Sequence
DNA	Deoxyribonucleic acid
bp	Base pair
CRL	Complex Resistance Locus
CDS	Calibrated Dichotomous Susceptibility
ESBL	Extended-spectrum beta-lactamase
ESC	Extended-spectrum cephalosporin
AGP	Antimicrobial Growth Promoter
ETEC	Enterotoxigenic *Escherichia coli*
IPEC	Intestinal Pathogenic *Escherichia coli*
APEC	Avian Pathogenic *Escherichia coli*
ExPEC	Extraintestinal Pathogenic *Escherichia coli*
EHEC	Enterohaemolytic *Escherichia coli*
STEC	Shiga-toxigenic *Escherichia coli*
EPEC	Enteropathogenic *Escherichia coli*
aEPEC	Atypical Enteropathogenic *Escherichia coli*
HGT	Horizontal Gene Transfer
MDR	Multiple-Drug (Antibiotic) Resistance
MLST	Multi-locus Sequence Typing
ST	Sequence Type
SNP	Single Nucleotide Polymorphism
WGS	Whole Genome Sequence/Sequencing

Abstract

Swine production is one of the largest agricultural industries in the world. The use of antimicrobials for treatment of disease outbreaks and as growth promoters drives the evolution of antimicrobial resistance in commensal populations of *Escherichia coli* in the pig gut. Therefore, the billions of tonnes of pig faeces generated by production every year are a serious environmental contaminant. Despite the status of *E. coli* as an important commensal and pathogen of both animals and humans, little is known about the genomic characteristics of porcine commensal *E. coli* in Australia. The clonal groups, antimicrobial resistance genes (ARGs), mobile genetic elements (MGEs) and virulence-associated genes (VAGs) they harbour remain poorly characterised.

In order to address this, we characterised 103 multidrug resistant commensal *E. coli* from two Australian farms with a history of antimicrobial use using whole genome sequencing. Phylogroup A and clonal complex 10 strains with global origins and a variety of virulence genotypes associated with extra-intestinal pathogenic *E. coli* (ExPEC) dominated the collection. Multiple class 1 integrons augmented by IS*26,* an important driver of evolution in antimicrobial resistance loci were observed in multiple sequence types. Whilst no resistance to critically important human antimicrobials was observed, this suggests swine production is a reservoir for the evolution of novel drug resistance characteristics that may transfer to human populations and rapidly acquire resistance to critically important antimicrobials.

We also identified an ST131 strain carrying an IncHI2 plasmid with multiple ARGs and a ColV plasmid conferring virulence traits. This strain was highly related to an ST131 strain isolated from a human infection. Remarkably, the human strain carried near identical plasmids. The apparent presence of the IncHI2 plasmid in multiple porcine strains, carriage of metal resistance and identical integrons strongly suggests the human pathogen originated from swine production.

Whilst there is a need to increase the number of porcine commensal *E. coli* whole genome sequences from multiple farms and states in Australia, our findings

support swine production as a reservoir of multiple drug resistance determinants. Furthermore, a proportion of porcine commensal *E. coli* are known or potential human pathogens. Genomic surveillance in swine and other intensively reared food production animals is critical to the ongoing management of the global issue of antimicrobial resistance.

Chapter 1: Thesis Overview

1.1 Overview

This thesis by compilation utilised whole genome sequencing to characterise 103 multidrug resistant *E. coli* isolated from the faeces of piglets on two Australian farms with an extensive history of antimicrobial use for the treatment of diarrhoeal disease. The first two results chapters were published during the course of the candidature and the final results chapter is being prepared for journal submission.

1.2 Aims

The three major aims of this thesis were to:

1. Determine the population structure, multi-locus sequence types, antimicrobial resistance genes, virulence-associated genes and plasmid types of a collection of multidrug resistant porcine commensal *E. coli*
2. Identify mobilised antimicrobial resistance genes, plasmids, clonal groups and strains and that may pose a risk to human and animal health
3. Use long read sequencing to fully characterise strains and plasmids identified in Aim 2 and compare them to relevant human origin strains

1.3 Summary and knowledge added to the field

This thesis addressed an absence of Australian porcine commensal *E. coli* whole genome sequences in the literature. It determined the presence of antimicrobial resistance genes, mobile genetic elements, virulence-associated genes as well as population structure and multilocus sequence types (STs) present.

The major findings included:

- The STs present in multidrug resistant porcine commensal *E. coli,* noting the dominance of phylogroup A and clonal complex 10
- Extensive genotypic drug resistance, a notable feature of which was the carriage of class 1 integrons associated with IS*26*
- The global origins of porcine CC10 *E. coli,* some of which are potential human pathogens

- Identification of an ST131 strain, a known human pathogen, carrying a large resistance plasmid that likely evolved within the context of swine production
- Evidence of potential zoonotic transfer of this ST131 strain and its plasmids in a case of human infection

Chapter 2: Literature Review

2.1 Antimicrobial Resistance: A Global One Health Issue

Antimicrobial resistance (AMR) is globally recognised as one of the greatest threats to human health this century. The 2014 World Health Organisation and 2016 UK Government reports on antimicrobial resistance have brought this issue to the forefront of human health awareness[1, 2]. The prospect of 10 million deaths and $100 trillion in cumulative costs by the year 2050 stands as a truly disturbing possibility. The potential costs in Australia are yet to be estimated, however the 2016 Antimicrobial Use and Resistance in Australia (AURA) report highlighted the issue as a priority for both the departments of Health and Agriculture[3].

Antimicrobial resistance arises in bacteria in response to the selective pressure of antimicrobial use in humans, agricultural systems and contamination of various environments with antimicrobial residues. It is therefore not simply a human health issue, but an animal and environmental health issue as well. The paradigm of One Health; the interconnectedness of human, animal and environmental health is exemplified by the issue of antimicrobial resistance and the problem needs to be tackled within this framework[4].

The vast majority of antimicrobial resistance in bacteria is mediated by specific genes that confer resistance by removing antimicrobial compounds from the cell or by modifying and degrading them so they are no longer efficacious. The dissemination of these genes and the bacteria that carry them is the fundamental basis of the problem of antimicrobial resistance (Fig. 2.1).

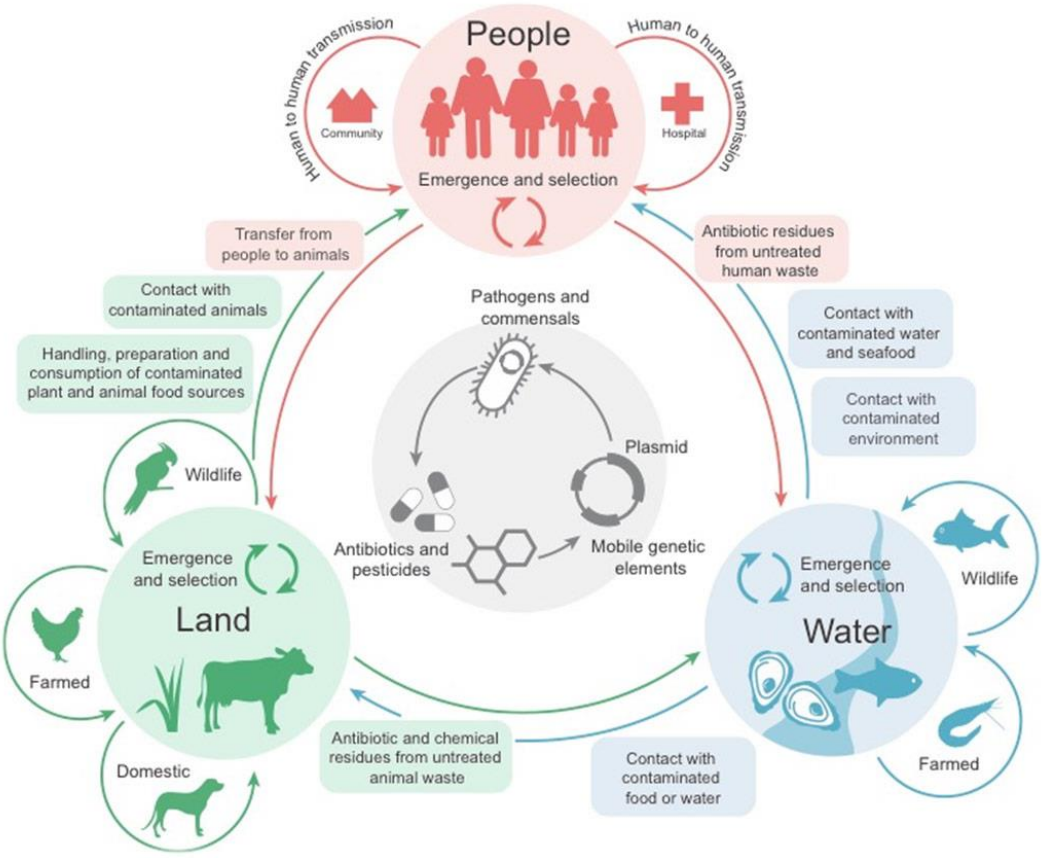


Figure 2.1

Schematic representation of the emergence and selection of drug resistant bacteria and MGEs within the One Health framework.

Three concepts that are central to understanding this dissemination are:

1. Clonal dissemination: The movement and proliferation of individual strains of bacteria that possess antimicrobial resistance genes (also known as 'vertical gene transfer').
2. Horizontal Gene Transfer (HGT): The transfer of antimicrobial resistance genes between different bacteria via MGEs.
3. Antimicrobial Selection Pressure: The selection of resistant bacteria due to the use of antimicrobials in human and agriculture as well as their presence in the wider environment.

Horizontal gene transfer is perhaps the most important component of antimicrobial resistance as it allows bacteria to rapidly spread resistance traits and promotes their continual evolution.

2.2 *E. coli*: the highly adaptable global barometer of MDR

E. coli is an ideal model organism for the study of antimicrobial resistance as it is near ubiquitous in humans, animals and the environment, it may be both a harmless commensal and a dangerous pathogen, and it can express resistance to every known antimicrobial class. As a commensal, its major habitat is the mammalian gut, whilst as a pathogen it may be split into two major subtypes[5]. Intestinal pathogenic *E. coli* (IPEC) cause disease within the gut of humans and animals, mediated by specific, generally well-characterised virulence factors[6]. Contrastingly, extra-intestinal pathogenic *E. coli* (ExPEC) infect non-intestinal sites such as the urinary tract, meninges, wounds and blood[7-9]. Most ExPEC are not diagnostically distinct from commensal strains and their epidemiology is still poorly understood. This issue is a major theme within this thesis and is discussed in greater depth below.

The ability of *E. coli* to pursue vastly differing lifestyles is due to metabolic and regulatory capabilities as well as genome plasticity[10]. Genome plasticity refers to the ability of an organism to gain and lose DNA by horizontal transfer, recombination or mutation in response to selective pressure such as antimicrobial use or other environmental factors. *E. coli* is particularly adept in this respect. As an organism that typically encodes about 5000 genes, the species pan genome is estimated to comprise 18,000 orthologous genes with as few as 2000 of these genes being considered 'core', or present in all strains[11].

Antimicrobial resistance in pathogenic lineages of *E. coli* increases morbidity and mortality as well as healthcare costs associated with their treatment[12]. The relative contribution of humans, animals and abiotic reservoirs to this problem is still poorly understood. Nonetheless, it is apparent that multidrug resistance has increased over time. A 2012 study of food and animal *E. coli* from the 1950 to 2002

indicated the proportion of MDR strains examined over this time period increased from 7.2% to 63.6%[13]. It is likely current rates are now even higher[12]. The emergence of pan-resistant strains, those that are resistant to every class of available antimicrobial is a major concern and dissemination of these strains into the wider ecosystem would be disastrous. Their presence in hospital sewage and wastewater is a public health issue that needs to be urgently addressed[14].

2.3 Does agriculture contribute to AMR in humans?

E. coli and AMR in food animals

E. coli in the gut of food production animals is recognized as a large reservoir or sink in which extensive intra- and inter-species HGT can occur[15]. The sharing of genetic information between pathogenic and commensal isolates may occur in this niche and may be accelerated when antimicrobial selective pressure is present[16].

Antimicrobial resistance is present in commensal isolates for two main reasons. Firstly, the antibiotics utilised to treat pathogenic infections places selective pressure on commensal as well as pathogenic bacteria, killing susceptible clones and allowing resistant clones to re-colonise the gut in higher numbers[17]. Secondly, the use of sub-therapeutic concentrations of antibiotics, known as antimicrobial growth promoters (AGPs), in-feed to reduce the occurrence of infection and promote healthy development, similarly places commensal *E. coli* in the gut under selective pressure[18]. Fortunately, Australia does not have the same issues with AGPs as other countries due to stricter regulations and the fact many antimicrobials used in animals overseas were never approved for animal use here. As a consequence of sound stewardship practices in Australia, there is only limited evidence of Gram-negative bacteria carrying genes encoding resistance to clinically important antimicrobials in pigs[19-21].

AMR in pig production

Pig production is a global business and the health of livestock is paramount for efficient feed conversion and delivery of safe pork products to consumers. Pigs are

a key primary industry for Australia. The Australian Pork Limited 2016-2017 annual report counted 5.16 million pigs at slaughter, yielding 397,000 metric tonnes of pork and a gross value of $1.277 billion[22]. This pales in comparison to China, which produces and consumes over half the world's pork on an annual basis with almost 55 million metric tonnes of pork predicted to be produced in 2018, representing an increase of 15 times over the past 50 years[23, 24]. Globally, pork consumption is on the rise and it is therefore important from a public health perspective to understand the AMR characteristics of potentially zoonotic organisms such as *E. coli* that are a significant by-product of such large scale production.

Pigs reared in intensive animal production systems are vulnerable to a wide variety of bacterial, viral and parasitic infectious threats. *E. coli* is significant in this regard, being the second most studied porcine pathogen by publication count[25]. Enterotoxigenic *E. coli* (ETEC) that cause gastrointestinal disease during neonatal, pre- and post-weaning stages are particularly significant. An outbreak of ETEC infection can lead to significant stock losses and is expensive to treat and control. Antimicrobials must be used to control gastrointestinal infections and as a result multi-drug resistant ETEC are common, especially on farms where recurrent ETEC infections occur[26]. Whilst ETEC are usually swine-specific due to the adhesins they encode, *E. coli* is increasingly reported as an extra-intestinal pathogen in swine[27-29]. This increase may have implications for the emergence of zoonotic pathogens among swine and remains a poorly explored area of study. The major issue however is that antimicrobial use in the treatment of any infection selects for multi-drug resistant commensal *E. coli* in the gut of treated animals, which increases the risk of dissemination of antimicrobial resistance genes into the biosphere[17, 30]. The potential scale of this spread is alarming given the ever-increasing scale of global production and the fact that pigs each produce around 5kg of liquid faeces a day, carrying between 10^4 and 10^8 *E. coli* per gram[23, 31].

Specific antimicrobial use can be correlated with resistance phenotypes and genotypes when usage data is available[32], however national and global antimicrobial usage is very difficult to monitor due to varying surveillance

methodologies utilised by reporting bodies in different countries as well as off-label use of antibiotics[18]. Antimicrobial use in the swine industry in Australia was last surveyed in 2009. 197 herds were surveyed accounting for roughly half the country's large pig herds (>200 sows)[33]. The study examined antibiotic use through injection, in-feed and in-water administration for the treatment of various common pig diseases such as ETEC infection. The most common antimicrobials used included penicillins, tetracycline, sulfonamides, macrolides, and apramycin/neomycin. The majority of these commonly used compounds are considered to be of low importance for human health by the World Health Organisation[34, 35] and the Australian Antimicrobial Resistance Standing Committee[36]. The accuracy of that assessment is debatable due to the potential for co-selection of resistance determinants[37, 38]. Resistance genes for both low importance and critically important antibiotics may be proximally located on the genome of an organism or on MGEs. This could allow for the selection of resistance to a critically important drug by treatment with a supposedly low importance drug.

Globally, resistance to ampicillin, streptomycin, sulfonamides and tetracycline is widely disseminated in commensal isolates from pigs and other food production animals[15, 39]. Multiple resistance to all four classes of these 'older' antibiotics that have been in use for decades is also common in commensal *E. coli* in Europe according to consecutive European Union reports[40, 41]. A 2011 investigation into resistance in pigs and poultry in China found high rates of multi-resistance; 81% in a pool of over 500 *E. coli* samples. Consistent with European reports, resistance to ampicillin, streptomycin, sulfonamides and tetracycline were most frequently observed in the Chinese study[42]. Within the food production chain, this resistance pattern persists as highlighted by another European study of *E. coli* in retail meat products. Forty-six *E. coli* isolates from retail pork were tested with 32.6% resistant to more than one class of antimicrobial and 8.7% resistant to more than four classes. Once again, resistance to ampicillin, sulfonamides, streptomycin and tetracycline were most frequently observed[43]. Few studies account for genotype to support phenotypic data however Szmolka *et al* found the dominant genes accounting for resistance phenotypes in pigs were *bla*$_{\text{TEM}}$, *sul2, aadA2*-like,

strA/*strB* and *tet*(A/B). Whilst there is very little data on commensal pig isolates from Australia, resistance phenotypes and genotypes of ETEC from pigs are highly similar to previously mentioned studies on commensals[44, 45]. Full genotypic characterisation of populations is absent and may be filled by generating large-scale genomic datasets that illustrate the situation for pigs in Australia.

Resistance to critically important human antibiotics including fluoroquinolones and extended-spectrum cephalosporins (ESCs) was recently reported within pathogenic and commensal *E. coli* of porcine origin in Australia[20, 46]. Whilst one farm had high levels of ESC resistance, the overall carriage of fluoroquinolone and ESC resistance in pathogenic strains from veterinary diagnostic labs was low. Their emergence in swine is concerning as fluoroquinolones are banned for use in food animals, indicating this resistance has infiltrated pig production from an external source. ESCs are a last-line, off-label treatment available for ETEC so the emergence of resistance in swine is a major concern[20]. Resistance to quinolones and extended-spectrum beta-lactams has also been observed in commensal isolates in other parts of the world[47-49]. In China, high levels of resistance to quinolones and cephalosporins in faecal *E. coli* from pigs highlights the potential for bacteria to acquire resistance to newer antimicrobials that play a significant role in human medicine if their use is not adequately managed and controlled[42].

The role of pig production in the evolution of AMR

Historically, studies of resistance in animal-associated *E. coli* have documented similarities to human strains with regard to resistance and virulence traits however these are limited by the fact that genes are usually screened for in isolation[50-54]. More recently, studies utilizing whole genome sequencing have increased the resolution of the comparisons that can be made, however numerous limitations are present in every study, including the exclusive selection of ESBL-producing *E. coli,* lack of temporally and geographically related isolates and unbalanced sample sizes from different reservoirs[55, 56]. Furthermore, comparisons with poultry and their associated meat products are more common than studies concerning other animals.

Whilst transfer between animals and humans undoubtedly occurs in both directions, studies have not addressed the evolution of resistance plasmids and other ARGs in association with MGEs within swine production. Many studies are eager to assign a source to the resistance genes or plasmids they find but ignore the potential role animal production could play in the evolution and amplification of successful plasmids and MGEs. This is critical, as resistance plasmids, regardless of origin, continue to evolve in these reservoirs under unique selective pressures and can be trafficked into human populations via multiple pathways. Whole genome sequence analysis of multiple plasmids is required to determine if resistance traits observed in clinical pathogens actually evolved or originated in animals or whether they have alternative origins. It is undoubted that both food animal and alternative sources play a role, though the relative contribution of each should be established.

In terms of swine production, antimicrobial use is the obvious factor that might drive the evolution of ARGs and MGEs that circulate in the environment, however feed supplementation with heavy metals is another major factor in swine production that is potentially overlooked[57-59]. Copper and zinc are common additives to pig feed and select commensal bacteria that are resistant to these heavy metals. These heavy metal resistance determinants are frequently found to be co-resident on plasmids that also carry ARGs. This linkage allows the continued selection of AMR plasmids in swine production even in the absence of antibiotic use and could be a potential indicator of plasmids that originate from swine and other production animals[60].

2.4 Genetic basis of AMR

Background

ARGs do not exist in isolation within bacterial genomes. In *E. coli* they tend to cluster together on mobile plasmids where minimal fitness cost is incurred by their acquisition, because critical metabolic genes are not affected and it is advantageous to possess resistance determinants[61]. These regions are known as

complex resistance loci (CRL) and may comprise ARGs, heavy metal resistance genes and virulence-associated genes in conjunction with a variety of MGEs responsible for their mobility. A number of MGEs such as plasmids, integrons, insertion sequences (IS) and transposons play a critical role in the mobility and dissemination of resistance genes in *E. coli*[62, 63]. Complex resistance loci are primarily generated by homologous recombination, site-specific recombination and the transposition of insertion sequences and transposons harbouring resistance genes[62]. These loci form incrementally, by a series of one-off genetic events that are selected for by antimicrobial use in food animals, clinics and residues in the environment. This yields both mosaic and nested genetic structure where individual or multiple modules may be individually mobile. A good example of this is the R100 plasmid, which carries multiple ARGs, transposons and IS elements on a self-transmissible plasmid backbone (Fig. 2.2)[64]. Despite their prevalence in commensal isolates, MGEs in pathogenic isolates are described far more frequently. Furthermore, most epidemiological studies only screen for presence of individual genes using PCR as opposed to characterizing the variable and mosaic structure of plasmids and resistance regions, which evolve in response to fluctuating selective pressures. These complex resistance structures persist because the use of one antimicrobial can select for physically associated genes mediating resistance to an array of antimicrobials. Characterising complex resistance loci and the plasmids that carry them is important because the MGEs that comprise part of their structure are capable of transferring their multiple resistance genes between bacteria in a single event allowing the rapid dissemination of MDR in bacterial communities.

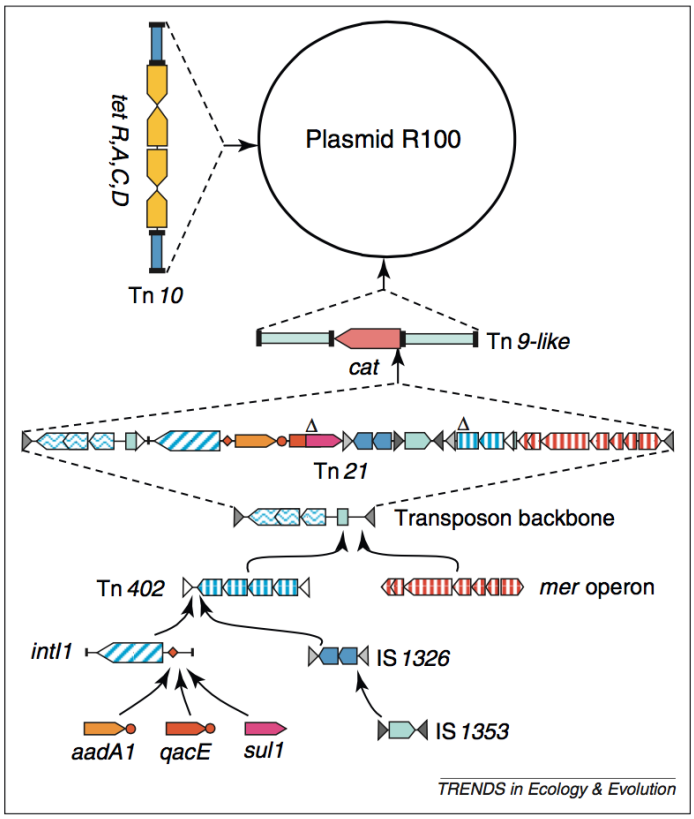


Figure 2.2

Schematic representation of R100, illustrating mosaic and nested structure of a complex resistance locus on a mobile plasmid backbone[64].

Plasmids

Plasmids are circular, self-replicating, extra-chromosomal DNA molecules that are carried by bacteria as accessory elements, typically encoding mechanisms of antimicrobial resistance, heavy metal resistance, virulence and fitness[65]. Plasmids are typed into incompatibility groups based on the sequence of their replication genes. Mechanisms of incompatibility ensure plasmids of the same incompatibility group do not co-reside in a single cell, insuring against the fitness cost associated with carrying redundant genetic material. Plasmid multilocus sequence typing (pMLST) schemes are available for a number of common incompatibility groups allowing greater resolution of lineages from an epidemiological perspective[66]. These schemes are useful yet still limited and whole genome sequencing and analysis of complete plasmids is preferable. Multidrug resistance plasmids in pathogenic *E. coli* are common and are increasingly characterised in their entirety as next-generation sequencing

approaches become mainstream, although closing plasmid sequences is still problematic using current short-read sequencing technologies. The emergence of long read sequencing is beneficial in this regard as read lengths can span across large CRLs and allow important plasmids to be fully characterised. Cost is still an issue for accessibility to this technology, however the number of closed plasmid sequences on GenBank generated by long read sequencing has increased dramatically over the past few years.

Plasmids that carry mosaic resistance regions typically form as a result of recombination between different plasmids, insertion of transposons and insertion sequences[15]. Incompatibility types carrying multiple resistance determinants in *E. coli* include IncF, IncI1, IncHI1, IncN and IncA/C (though IncA/C has since been shown to comprise separate and compatible IncA and IncC groups)[65, 67-73]. These plasmid types are reported globally carrying ARGs in pigs however literature has recently become extremely biased towards plasmids carrying extended spectrum beta-lactamases (ESBLs) and colistin ARGs such as *bla*$_{\text{CTX-M}}$ and *mcr* variants[74-78]. The use of these antimicrobials in overseas food production systems is a significant factor in the emergence of resistance[75]. Plasmid-mediated colistin resistance has not been reported in Australian pigs, likely due to the fact it has never been used in Australian food production. However, *bla*$_{\text{CTX-M}}$-carrying IncI1 plasmids were recently identified in an Australian pig farm with a history of ceftiofur use[20]. It is evident that specific antimicrobial use dictates the abundance of plasmids carrying resistance determinants in pig production. However, it is important to determine if these plasmids are simply from external sources in response to selective pressure or whether plasmids that are already successful in pigs acquire the determinants on other mobile elements. Both of these scenarios occur, though the balance between them depends on a myriad of factors with trends likely to be different on local, regional and global scales. Other plasmid types previously reported in Australian pigs include IncF and IncHI1, though it is apparent significant diversity exists[79]. Whole genome sequence analysis of plasmids and the ARGs they carry in porcine *E. coli* from Australia is currently lacking.

Class 1 integrons

Integrons are gene-capture and expression systems that are capable of both acquiring and losing gene cassettes. Integrons capture and express gene cassettes, typically conferring resistance to antimicrobials, which may be expressed in tandem; this is known as a cassette array. This allows bacteria to express an MDR phenotype with minimal fitness cost, as other genes are not interrupted by the insertion of additional cassettes[80, 81]. Gene cassettes are circularised promoter-less open reading frames with an attachment site *attC,* for site-specific recombination with the variable region of the integron (Fig. 2.3). Gene cassettes encode a wide range of proteins and are widespread in soil and marine environments[82]. While most well characterized gene cassettes are sourced from clinically relevant bacterial pathogens and encode resistance to mainstream antibiotics, the vast majority of gene cassettes from bacterial populations that reside in soil and aquatic environments encode functionally uncharacterised hypothetical proteins[83]. The abundance of uncharacterized genes reflects the evolutionary utility of integrons as highly variable gene expression systems allowing bacteria to quickly adapt to different environmental and anthropogenic pressures. This ability is invaluable to bacteria under the stress of antibiotic treatment and provides some explanation as to the abundance of integrons now observed in clinical bacteria. It is a major concern that these human-evolved integrons are now present in previously pristine natural environments as a form of anthropogenic pollution[84].

Integrons are designated classes based on the nucleotide sequence of the integrase gene, with classes 1, 2 and 3 being the most commonly observed in clinical and commensal *E. coli* isolates[83]. Bounded by inverted repeats, conserved or 'typical' class 1 integrons consist of two conserved segments that flank the cassette array known as the 5`-CS and 3`-CS. The 5`-CS consists of the class 1 integrase gene *intI1* encoding a site-specific recombinase responsible for insertion of free gene cassettes into the variable region at the *attI* site. This recombinase also contains a promoter region so that gene cassettes can be shuffled and upregulated in proximity to the promoter. The 3`-CS consists of *sul1* sulfonamide resistance gene, *qacEΔ1* and a truncated transposition module *tniAΔ* (Fig. 2.3)[85]. Clinically, they

are usually observed on plasmids of varying incompatibility types and genomic pathogenicity islands however within environmental isolates they are often located on the chromosome[63]. 'Atypical' class 1 integrons are those that do not exhibit the classical structure and may form by recombination or interruption by IS elements. Integrons that possess a *sul3* 3`-CS instead of *sul1* are one such example that have been identified in humans and animals globally[86-89]. IS interrupted integrons are also common and may be associated with composite transposons carrying multiple ARGs[90, 91]. Atypical integrons are important because their association with IS elements provides a platform for the evolution of novel resistance structures via IS-mediated transposition and recombination. IS insertions may also be used as genetic signatures to track the spread of specific integrons for epidemiological purposes[92].

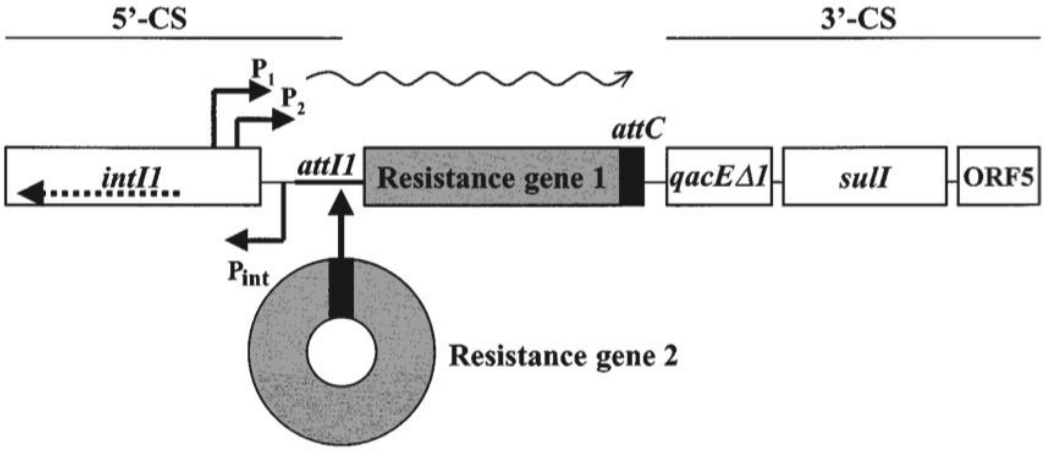


Figure 2.3.

Schematic diagram of class 1 integron showing 5` and 3` conserved segments, attI and attC recombination sites and gene cassettes[63].

As integrons are not individually mobile, their mobility is dependent on associated mobile elements such as plasmids and transposons such as mercury resistance transposon Tn*21*[93]. Tn*21* has played a significant role in the dissemination of antimicrobial resistance in association with class 1 integrons. The ubiquity of antimicrobials and mercury in the environment provide strong co-selective pressure for the maintenance of both Tn*21* and associated integrons.

Numerous international reports indicate integrons are widespread in swine, encoding resistance to antimicrobials commonly used in production, though most studies are limited to PCR data and none have performed whole genome sequencing to elucidate their diversity and contextualise their location on plasmids[39, 94-96]. Prior to this study there were no whole genome sequencing papers characterising integrons in Australian pigs and only one PCR study[44].

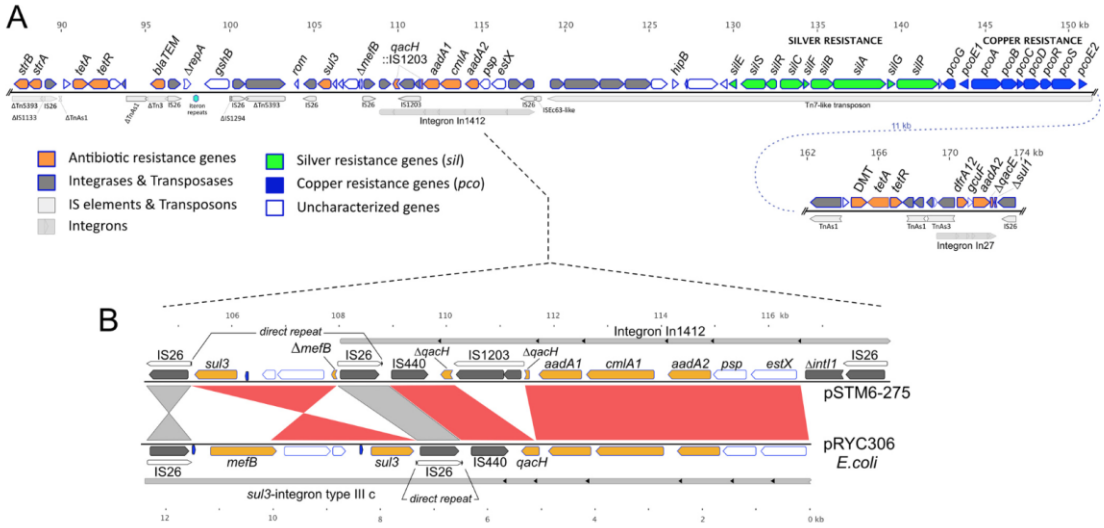
Insertion sequences

Insertion sequences (IS) are small MGEs typically defined by inverted repeats and an open reading frame encoding a transposase[97]. They persist in bacterial genomes due to their common association with antimicrobial resistance genes [97]. When an insertion is not fatal to an organism and selective pressure is present, bacteria are more likely to retain antimicrobial resistance gene-associated IS elements. This stable insertion site is then established as a location for the acquisition of further mobile elements and resistance genes which may form complex resistance loci as they accumulate[62]. IS are incredibly diverse and associated with a wide variety of ARGs in gram-negative bacteria[98]. IS*26* is arguably the most significant and successful of these elements for a number of reasons. It is associated with many different resistance genes, has a unique mechanism of transposition, recognises itself, has no copy number control, forms novel mobile CRLs, and plays a role in plasmid evolution, stability and formation of novel co-integrates[91, 97-106]. The carriage of IS elements in Australian porcine commensal *E. coli* is yet to be studied.

Transposons

Transposons are MGEs that encode mechanisms for their own excision and insertion into non-homologous DNA molecules. They carry accessory genes, which are often ARGs, in a wide range of hosts. Within *E. coli* common transposons include derivatives of Tn*3,* Tn*5* and Tn*7*, which encode resistance to ampicillin, kanamycin, and trimethoprim and streptomycin respectively[107]. They are typically observed on plasmids and within genomic islands in association with integrons and insertion sequences. Metal resistance transposons in particular are often associated with ARGs on plasmids due to environmental contamination with

heavy metals such as mercury, copper and zinc[60]. Wastewater is a particularly important reservoir of these plasmids though they may be selected in swine production due to in-feed copper supplementation[71, 108]. Tn*21* is a key example of this phenomenon due to the role it plays in mobilizing integrons in both natural and clinical environments [93]. A Tn*7*-like transposon is also known to carry copper and silver resistance genes in *E. coli* and *S. enterica* species including those



originating from pigs (Fig. 2.4) [108, 109].

Figure 2.4

Example of an atypical *sul3* integron, associated with Tn*7*-like copper/silver resistance transposon on an IncHI2 plasmid backbone from an Australian porcine *S. enterica* strain. Figure from Billman-Jacobe *et al*[109].

2.5 *E. coli* in humans and animals; where do ExPEC originate?

In addition to the evolution and spread of AMR via commensal *E. coli* that do not usually cause infectious disease, animal production poses a risk to human health via the evolution of zoonotic pathogens. Whilst the spread of AMR alone is a major concern, the evolution of MDR zoonotic pathogens increases the severity of the consequences. The case of O157:H7 enterohaemolytic *E. coli* (EHEC) originating from cattle represents a well-documented instance of MDR zoonosis[110]. By contrast, much debate still remains over the contribution of animals to ExPEC in humans. ExPEC are defined as *E. coli* causing disease outside the gastrointestinal

tract with subtypes including uropathogenic (UPEC), those that cause urinary tract infections, pyelonephritis and sepsis, and neonatal meningitis-associated (NMEC), those that cause meningitis in newborns. They represent a significant burden on healthcare worldwide and disproportionately affect women and immune-compromised individuals such as children and the elderly[111]. Drug resistant ExPEC exacerbate rates of morbidity and mortality and increase financial costs associated with treatment.

The evidence for human ExPEC being a case of zoonosis is less clear-cut than intestinal pathotypes, however there is a growing body of literature that suggests ExPEC zoonosis does occur. The difficulty in proving ExPEC zoonosis is mostly due to due to the lack of consensus regarding the conserved molecular characteristics of ExPEC. There is significant genetic diversity in *E. coli* that cause extra-intestinal infections because they are actually a subset of commensal *E. coli* resident in the faecal microbiome. Furthermore, unlike EHEC, which encode toxins that may be used as diagnostic markers, ExPEC exhibit significant variability with respect to the virulence-associated factors they carry and cannot be reliably classified using molecular markers despite attempts to do so[112, 113].

Due to these issues of classification, ExPEC is typically defined based on isolation site. Despite the lack of clear classification, numerous virulence-associated genes (VAGs) are commonly found within ExPEC. VAG is a useful term in this regard because the presence of these genes does not necessarily confirm pathogenicity. VAGs within ExPEC include adhesins, toxins, nutrient acquisition systems and protectins among others[7]. Many of these factors have been demonstrated to play a role in colonisation of the urinary tract, however many of them are also associated with intestinal fitness[114]. This suggests that ExPEC should be considered opportunistic pathogens.

The concept of ExPEC zoonosis requires that *E. coli* of animal faecal origin contaminating meat at slaughter may be consumed by a human and pass asymptomatically through the gastrointestinal tract before cross-contamination of the urogenital tract resulting in cystitis[51, 115]. Supporting evidence includes

shared VAGs and STs in animal, meat and human *E. coli* and the ability of food isolates to cause UTI in mice[92, 112, 116]. This model has the potential to explain one pathway for the transfer of ARGs and pathogens from agricultural sources to human pathogens. A recent genomic study provided the most convincing evidence of this pathway thus far, showing that a proportion of human urinary tract infections were likely originating in poultry meat in Flagstaff, Arizona[117]. However, it is certainly not the only potential pathway and the current data is not sufficient to infer how frequently this occurs. Furthermore, it is not reasonable to assume that the movement of drug resistant organisms between animals and humans is unidirectional. Other sources need to be investigated thoroughly including natural waterways and wastewater treatment facilities that release effluent contaminated with ARGs and pathogens back into the environment[64]. There are diverse pathways for bacteria and resistance genes to move between various niches and far more work needs to be done to understand the relative contribution of each to the issue of AMR in human infections[52].

Global pandemic lineages and the case for zoonosis

A number of pandemic lineages of ExPEC are responsible for a large proportion of extra-intestinal infections in humans. Within urinary tract and bloodstream infections these include ST131, ST69, ST95 and ST73 among other less prevalent clonal groups[114, 118]. ST131 is one of the most common and important sequence types isolated from urinary tract infections globally, primarily due to its resistance to fluoroquinolones and third generation cephalosporins[119]. The evolutionary and epidemiological history of this pandemic sequence type is still not fully understood, however a highly clonal sub-lineage, known as *H*30*Rx,* is globally distributed and carries conserved fluoroquinolone and cephalosporin resistance. It was initially thought that its success was primarily due to SNPs conferring fluoroquinolone resistance and carriage of $bla_{\text{CTX-M}}$ ESBL genes. However, the global distribution of its antimicrobial susceptible evolutionary precursors *H*22 and *H*41 has somewhat complicated a comprehensive understanding of ST131. *H*22 and *H*41 have been isolated from wild, domestic and food production animals and the reason for their success is likely to be multifactorial[117, 120, 121]. A recent study identified *H*22 as a potential zoonotic

agent being transferred from chickens to humans via meat[117]. ST95 is another sequence type that has been identified as potentially zoonotic due to its association with poultry, UTIs and meningitis[122]. ST95 strains typically don't express multi-resistance and are often fully susceptible, implying that drug resistance is not the only factor that plays a role in the emergence of pandemic clones[114, 123].

Whole genome analysis needs to be undertaken on spatially and temporally matched isolates from each niche within the pathway to determine which sequence types present a risk and the frequency at which these transfer events occur. This project will not go so far, however it will provide characterized genomes of porcine and human origin for future comparison with publicly available ExPEC genomes.

Pigs and ExPEC

Poultry are the most frequently implicated food animal that may play a role in the evolution and spread of drug resistant ExPEC to humans. By contrast, pigs are neither recognised as a source nor well studied in this regard. Various combinations of ExPEC VAGs have been observed in commensal *E. coli* from pigs in Australia however phylogenetic background was not comprehensively established[124]. A 2008 German study examined 331 *E. coli* isolates from a sow and five offspring pre- and post-weaning and found a correlation between higher numbers of VAGs and successful gut colonization in piglets, reiterating the association of VAGs with fitness[125]. The rise of ExPEC infections within pigs in China in particular raises the possibility that pigs may also be a source of ExPEC that could infect humans. Surprisingly, a number of these reports identify phylogroup A and clonal complex 10 strains as the most common groups causing infections[28]. Phylogroup is A is typically considered to be commensal and non-pathogenic, whilst CC10 is a common commensal of humans and animals. Furthermore, CC10 strains are increasingly reported as extra-intestinal pathogens. These studies typically fail to comprehensively compare virulence, AMR and phylogenetic characteristics together; instead relying on outdated phylogrouping methods and limited arrays of VAGs. Whole genome sequencing of larger collections that cover all these aspect are required to develop an understanding of

the potential role of swine production in the evolution of MDR ExPEC.

2.6 Conclusions

Comprehensive whole genome-based epidemiology of porcine *E. coli* is yet to be undertaken. Studies to date tend to compare virulence, AMR or phylogeny in isolation, which limits the epidemiological conclusions that can be made when comparing to spatially or temporally unrelated sequences. This thesis will contribute to this gap in the literature by characterising the population structure, sequence types, VAGs, ARGs, integrons, plasmids types and MGEs present in a collection of MDR porcine *E. coli* from Australia.

2.6 References

1. Organization WH. Antimicrobial resistance: global report on surveillance.2014.
2. O'Neill J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. London: Wellcome Trust, UK Government2016.
3. (ACSQHC) ACoSaQiHC. AURA 2016: first Australian report on antimicrobial use and resistance in human health.2016.
4. **Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E et al.** Tackling antibiotic resistance: the environmental framework. *Nat Rev Microbiol* 2015;13(5):310-317.
5. **Leimbach A, Hacker J, Dobrindt U**. E. coli as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity. In: Dobrindt U, Hacker JH, Svanborg C (editors). *Between Pathogenicity and Commensalism (Current Topics in Microbiology and Immunology*: Springer Berlin Heidelberg; 2013. pp. 3-32.
6. **Kaper JB, Nataro JP, Mobley HL**. Pathogenic Escherichia coli. *Nat Rev Microbiol* 2004;2(2):123-140.
7. **Johnson JR, Russo TA**. Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli. *Int J Med Microbiol* 2005;295(6-7):383-404.
8. **Petkovsek Z, Elersic K, Gubina M, Zgur-Bertok D, Starcic Erjavec M**. Virulence potential of Escherichia coli isolates from skin and soft tissue infections. *J Clin Microbiol* 2009;47(6):1811-1817.
9. **Ewers C, Li G, Wilking H, Kiessling S, Alt K et al.** Avian pathogenic, uropathogenic, and newborn meningitis-causing Escherichia coli: how closely related are they? *Int J Med Microbiol* 2007;297(3):163-176.
10. **Tenaillon O, Skurnik D, Picard B, Denamur E**. The population genetics of commensal Escherichia coli. *Nat Rev Microbiol* 2010;8(3):207-217.
11. **Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S et al.** Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet* 2009;5(1):e1000344.
12. **Vila J, Saez-Lopez E, Johnson JR, Romling U, Dobrindt U et al.** Escherichia coli: an old friend with new tidings. *FEMS Microbiol Rev* 2016;40(4):437-463.

13. **Tadesse DA, Zhao S, Tong E, Ayers S, Singh A et al.** Antimicrobial drug resistance in Escherichia coli from humans and food animals, United States, 1950-2002. *Emerg Infect Dis* 2012;18(5):741-749.
14. **Meir-Gruber L, Manor Y, Gefen-Halevi S, Hindiyeh MY, Mileguir F et al.** Population Screening Using Sewage Reveals Pan-Resistant Bacteria in Hospital and Community Samples. *PLoS One* 2016;11(10):e0164873.
15. **Szmolka A, Nagy B**. Multidrug resistant commensal Escherichia coli in animals and its impact for public health. *Front Microbiol* 2013;4:258.
16. **Baharoglu Z, Bikard D, Mazel D**. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet* 2010;6(10):e1001165.
17. **Holman DB, Chenier MR**. Temporal changes and the effect of subtherapeutic concentrations of antibiotics in the gut microbiota of swine. *FEMS Microbiol Ecol* 2014;90(3):599-608.
18. **Barton MD**. Impact of antibiotic use in the swine industry. *Current opinion in microbiology* 2014;19:9-15.
19. **Abraham S, Jordan D, Wong HS, Johnson JR, Toleman MA et al.** First detection of extended-spectrum cephalosporin- and fluoroquinolone-resistant Escherichia coli in Australian food-producing animals. *J Glob Antimicrob Resist* 2015;3(4):273-277.
20. **Abraham S, Kirkwood RN, Laird T, Saputra S, Mitchell T et al.** Dissemination and persistence of extended-spectrum cephalosporin-resistance encoding IncI1-blaCTXM-1 plasmid among Escherichia coli in pigs. *The ISME Journal* 2018;12(10):2352-2362.
21. **Reid CJ, Wyrsch ER, Roy Chowdhury P, Zingali T, Liu M et al.** Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 2017;3(12).
22. (APL) APL. Annual report 2016-20172017.
23. **Han S**. *Environmental Impacts of China's Pork Industry*. Wilson Center; 2014.
24. (USDA) USDoA. Livestock and Poultry: World Markets and Trade2018.
25. **VanderWaal K, Deen J**. Global trends in infectious diseases of swine. *Proceedings of the National Academy of Sciences* 2018;115(45):11495-11500.
26. **Chapman TA, Wu XY, Barchia I, Bettelheim KA, Driesen S et al.** Comparison of virulence gene profiles of Escherichia coli strains isolated from healthy and diarrheic swine. *Appl Environ Microbiol* 2006;72(7):4782-4795.
27. **Luppi A**. Swine enteric colibacillosis: diagnosis, therapy and antimicrobial resistance. *Porcine health management* 2017;3:16.
28. **Ding Y, Tang X, Lu P, Wu B, Xu Z et al.** Clonal analysis and virulent traits of pathogenic extraintestinal Escherichia coli isolates from swine in China. *BMC veterinary research* 2012;8:140.
29. **Tan C, Tang X, Zhang X, Ding Y, Zhao Z et al.** Serotypes and virulence genes of extraintestinal pathogenic Escherichia coli isolates from diseased pigs in China. *Veterinary journal (London, England : 1997)* 2012;192(3):483-488.
30. **Chantziaras I, Boyen F, Callens B, Dewulf J**. Correlation between veterinary antimicrobial use and antimicrobial resistance in food-producing animals: a report on seven countries. *J Antimicrob Chemother* 2014;69(3):827-834.
31. **Herrero-Fresno A, Larsen I, Olsen JE**. Genetic relatedness of commensal Escherichia coli from nursery pigs in intensive pig production in Denmark and

molecular characterization of genetically different strains. *J Appl Microbiol* 2015;119(2):342-353.
32. **Varga C, Rajic A, McFall ME, Reid-Smith RJ, Deckert AE et al.** Associations between reported on-farm antimicrobial use practices and observed antimicrobial resistance in generic fecal Escherichia coli isolated from Alberta finishing swine farms. *Prev Vet Med* 2009;88(3):185-192.
33. **Jordan D, Chin JJ, Fahy VA, Barton MD, Smith MG et al.** Antimicrobial use in the Australian pig industry: results of a national survey. *Australian veterinary journal* 2009;87(6):222-229.
34. **Organization WH**. Critically important antimicrobials for human medicine: categorization for the development of risk management strategies to contain antimicrobial resistance due to non-human antimicrobial use: report of the second WHO Expert Meeting, Copenhagen, 29-31 May 2007. 2007.
35. **Organization WH**. Critically important antibacterial agents for human medicine for risk management strategies of non-human use: report of a WHO working group consultation, 15-18 February 2005, Canberra, Australia. 2005.
36. Committee AARS. AMRSC Importance Ratings and Summary of Antibacterial Uses in Humans in Australia. Canberra2014.
37. **Bischoff KM, White DG, McDermott PF, Zhao S, Gaines S et al.** Characterization of Chloramphenicol Resistance in Beta-Hemolytic Escherichia coli Associated with Diarrhea in Neonatal Swine. *Journal of Clinical Microbiology* 2002;40(2):389-394.
38. **Makita K, Goto M, Ozawa M, Kawanishi M, Koike R et al.** Multivariable Analysis of the Association Between Antimicrobial Use and Antimicrobial Resistance in Escherichia coli Isolated from Apparently Healthy Pigs in Japan. *Microb Drug Resist* 2015.
39. **Lee M, Shin E, Lee Y**. Antimicrobial resistance and integron profiles in multidrug-resistant Escherichia coli isolates from pigs. *Foodborne Pathog Dis* 2014;11(12):988-997.
40. **EFSA**. EU Summary Report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2013. *EFSA Journal* 2015;13(2).
41. **EFSA**. EU Summary Report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2010. *EFSA Journal* 2012;10(3).
42. **Jiang HX, Lu DH, Chen ZL, Wang XM, Chen JR et al.** High prevalence and widespread distribution of multi-resistant Escherichia coli isolates in pigs and poultry in China. *Vet J* 2011;187(1):99-103.
43. **Kaesbohrer A, Schroeter A, Tenhagen BA, Alt K, Guerra B et al.** Emerging antimicrobial resistance in commensal Escherichia coli with public health relevance. *Zoonoses Public Health* 2012;59 Suppl 2:158-165.
44. **Abraham S, Trott DJ, Jordan D, Gordon DM, Groves MD et al.** Phylogenetic and molecular insights into the evolution of multidrug-resistant porcine enterotoxigenic Escherichia coli in Australia. *Int J Antimicrob Agents* 2014;44(2):105-111.
45. **Smith MG, Jordan D, Gibson JS, Cobbold RN, Chapman TA et al.** Phenotypic and genotypic profiling of antimicrobial resistance in enteric Escherichia coli communities isolated from finisher pigs in Australia. *Aust Vet J* 2016;94(10):371-376.

46. **Abraham S, Jordan D, Wong HS, Johnson JR, Toleman MA 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(4):273-277.
47. **Szmolka A, Anjum MF, La Ragione RM, Kaszanyitzky EJ, Nagy B**. Microarray based comparative genotyping of gentamicin resistant Escherichia coli strains from food animals and humans. *Veterinary microbiology*, Research Support, Non-U.S. Gov't 2012;156(1-2):110-118.
48. **Changkaew K, Intarapuk A, Utrarachkij F, Nakajima C, Suthienkul O et al.** Antimicrobial Resistance, Extended-Spectrum beta-Lactamase Productivity, and Class 1 Integrons in Escherichia coli from Healthy Swine. *Journal of food protection* 2015;78(8):1442-1450.
49. **Fortini D, Fashae K, Garcia-Fernandez A, Villa L, Carattoli A**. Plasmid-mediated quinolone resistance and beta-lactamases in Escherichia coli from healthy animals from Nigeria. *J Antimicrob Chemother* 2011;66(6):1269-1272.
50. **Jakobsen L, Garneau P, Bruant G, Harel J, Olsen SS et al.** 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.
51. **Nordstrom L, Liu CM, Price LB**. Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. *Front Microbiol* 2013;4:29.
52. **Singer RS**. Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. *Front Microbiol* 2015;6:28.
53. **Kluytmans JA, Overdevest IT, Willemsen I, Kluytmans-van den Bergh MF, van der Zwaluw K et al.** Extended-spectrum beta-lactamase-producing Escherichia coli from retail chicken meat and humans: comparison of strains, plasmids, resistance genes, and virulence factors. *Clin Infect Dis* 2013;56(4):478-487.
54. **Yamaji R, Friedman CR, Rubin J, Suh J, Thys E et al.** A Population-Based Surveillance Study of Shared Genotypes of Escherichia coli Isolates from Retail Meat and Suspected Cases of Urinary Tract Infections. *mSphere* 2018;3(4).
55. **Dorado-Garcia A, Smid JH, van Pelt W, Bonten MJM, Fluit AC et al.** Molecular relatedness of ESBL/AmpC-producing Escherichia coli from humans, animals, food and the environment: a pooled analysis. *J Antimicrob Chemother* 2018;73(2):339-347.
56. **Pietsch M, Irrgang A, Roschanski N, Brenner Michael G, Hamprecht A et al.** Whole genome analyses of CMY-2-producing Escherichia coli isolates from humans, animals and food in Germany. *BMC Genomics* 2018;19(1):601.
57. **Fard RM, Heuzenroeder MW, Barton MD**. Antimicrobial and heavy metal resistance in commensal enterococci isolated from pigs. *Veterinary microbiology* 2011;148(2-4):276-282.
58. **Holman DB, Chenier MR**. Antimicrobial use in swine production and its effect on the swine gut microbiota and antimicrobial resistance. *Canadian journal of microbiology* 2015;61(11):785-798.
59. **Holzel CS, Muller C, Harms KS, Mikolajewski S, Schafer S et al.** Heavy metals in liquid pig manure in light of bacterial antimicrobial resistance. *Environ Res*, Research Support, Non-U.S. Gov't 2012;113:21-27.
60. **Bednorz C, Oelgeschlager K, Kinnemann B, Hartmann S, Neumann K et al.** The broader context of antibiotic resistance: zinc feed supplementation of piglets increases

the proportion of multi-resistant Escherichia coli in vivo. *Int J Med Microbiol* 2013;303(6-7):396-403.
61. **Walsh C, Fanning S**. Antimicrobial resistance in foodborne pathogens--a cause for concern? *Current drug targets* 2008;9(9):808-815.
62. **Stokes HW, Gillings MR**. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol Rev* 2011;35(5):790-819.
63. **Carattoli A**. Importance of integrons in the diffusion of resistance. *Vet Res* 2001;32(3-4):243-259.
64. **Gillings MR, Stokes HW**. Are humans increasing bacterial evolvability? *Trends in Ecology & Evolution* 2012;27(6):346-352.
65. **Fernandez-Alarcon C, Singer RS, Johnson TJ**. Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic Escherichia coli from multiple animal sources. *PLoS One* 2011;6(8):e23415.
66. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
67. **Cain AK, Hall RM**. Evolution of a multiple antibiotic resistance region in IncHI1 plasmids: reshaping resistance regions in situ. *J Antimicrob Chemother*, Research Support, Non-U.S. Gov't 2012;67(12):2848-2853.
68. **Cain AK, Liu X, Djordjevic SP, Hall RM**. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant Salmonella enterica serovar Typhimurium from Australian animals. *Microbial drug resistance (Larchmont, NY)* 2010;16(3):197-202.
69. **Szczepanowski R, Braun S, Riedel V, Schneiker S, Krahn I et al.** The 120 592 bp IncF plasmid pRSB107 isolated from a sewage-treatment plant encodes nine different antibiotic-resistance determinants, two iron-acquisition systems and other putative virulence-associated functions. *Microbiology* 2005;151(Pt 4):1095-1111.
70. **Venturini C, Beatson SA, Djordjevic SP, Walker MJ**. Multiple antibiotic resistance gene recruitment onto the enterohemorrhagic Escherichia coli virulence plasmid. *FASEB J* 2010;24(4):1160-1166.
71. **Wibberg D, Szczepanowski R, Eikmeyer F, Puhler A, Schluter A**. The IncF plasmid pRSB225 isolated from a municipal wastewater treatment plant's on-site preflooder combining antibiotic resistance and putative virulence functions is highly related to virulence plasmids identified in pathogenic E. coli isolates. *Plasmid* 2013;69(2):127-137.
72. **Carattoli A**. Resistance plasmid families in Enterobacteriaceae. *Antimicrobial agents and chemotherapy* 2009;53(6):2227-2238.
73. **Ambrose SJ, Harmer CJ, Hall RM**. Compatibility and entry exclusion of IncA and IncC plasmids revisited: IncA and IncC plasmids are compatible. *Plasmid* 2018;96-97:7-12.
74. **de Been M, Lanza VF, de Toro M, Scharringa J, Dohmen W et al.** Dissemination of cephalosporin resistance genes between Escherichia coli strains from farm animals and humans by specific plasmid lineages. *PLoS Genet* 2014;10(12):e1004776.
75. **Garcia-Menino I, Garcia V, Mora A, Diaz-Jimenez D, Flament-Simon SC et al.** Swine Enteric Colibacillosis in Spain: Pathogenic Potential of mcr-1 ST10 and ST131 E. coli Isolates. *Front Microbiol* 2018;9:2659.

76. **Li R, Xie M, Zhang J, Yang Z, Liu L et al.** Genetic characterization of mcr-1-bearing plasmids to depict molecular mechanisms underlying dissemination of the colistin resistance determinant. *J Antimicrob Chemother* 2017;72(2):393-401.
77. **Lindsey RL, Frye JG, Thitaram SN, Meinersmann RJ, Fedorka-Cray PJ et al.** Characterization of multidrug-resistant Escherichia coli by antimicrobial resistance profiles, plasmid replicon typing, and pulsed-field gel electrophoresis. *Microb Drug Resist* 2011;17(2):157-163.
78. **Pulss S, Semmler T, Prenger-Berninghoff E, Bauerfeind R, Ewers C**. First report of an Escherichia coli strain from swine carrying an OXA-181 carbapenemase and the colistin resistance determinant MCR-1. *International Journal of Antimicrobial Agents* 2017;50(2):232-236.
79. **Abraham S, Gordon DM, Chin J, Brouwers HJ, Njuguna P et al.** Molecular characterization of commensal Escherichia coli adapted to different compartments of the porcine gastrointestinal tract. *Appl Environ Microbiol* 2012;78(19):6799-6803.
80. **Labbate M, Boucher Y, Luu I, Chowdhury PR, Stokes HW**. Integron associated mobile genes: Just a collection of plug in apps or essential components of cell network hardware? *Mob Genet Elements* 2012;2(1):13-18.
81. **Guardabassi L, Courvalin P**. Modes of Antimicrobial Action and Mechanisms of Bacterial Resistance. *Antimicrobial Resistance in Bacteria of Animal Origin*: American Society of Microbiology; 2006.
82. **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.
83. **Gillings MR**. Integrons: past, present, and future. *Microbiol Mol Biol Rev* 2014;78(2):257-277.
84. **Power ML, Samuel A, Smith JJ, Stark JS, Gillings MR et al.** Escherichia coli out in the cold: Dissemination of human-derived bacteria into the Antarctic microbiome. *Environmental Pollution* 2016;215:58-65.
85. **Stokes HW, Hall RM**. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* 1989;3(12):1669-1683.
86. **Antunes P, Machado J, Peixe L**. Dissemination of sul3-containing elements linked to class 1 integrons with an unusual 3' conserved sequence region among Salmonella isolates. *Antimicrob Agents Chemother* 2007;51(4):1545-1548.
87. **Curiao T, Canton R, Garcillan-Barcia MP, de la Cruz F, Baquero F et al.** Association of composite IS26-sul3 elements with highly transmissible IncI1 plasmids in extended-spectrum-beta-lactamase-producing Escherichia coli clones from humans. *Antimicrob Agents Chemother* 2011;55(5):2451-2457.
88. **Liu J, Keelan P, Bennett PM, Enne VI**. Characterization of a novel macrolide efflux gene, mef(B), found linked to sul3 in porcine Escherichia coli. *J Antimicrob Chemother* 2009;63(3):423-426.
89. **Moran RA, Holt KE, Hall RM**. pCERC3 from a commensal ST95 Escherichia coli: A ColV virulence-multiresistance plasmid carrying a sul3-associated class 1 integron. *Plasmid* 2016.
90. **Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP et al.** Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in Escherichia coli of human and animal origin. *PLoS One* 2010;5(9):e12754.

91. **Reid CJ, Roy Chowdhury P, Djordjevic SP**. Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. *Plasmid* 2015;80:127-137.
92. **Manges AR, Harel J, Masson L, Edens TJ, Portt A et al.** Multilocus sequence typing and virulence gene profiles associated with Escherichia coli from human and animal sources. *Foodborne Pathog Dis* 2015;12(4):302-310.
93. **Liebert CA, Hall RM, Summers AO**. Transposon Tn21, flagship of the floating genome. *Microbiology and molecular biology reviews : MMBR* 1999;63(3):507-522.
94. **Lay KK, Koowattananukul C, Chansong N, Chuanchuen R**. Antimicrobial resistance, virulence, and phylogenetic characteristics of Escherichia coli isolates from clinically healthy swine. *Foodborne Pathog Dis* 2012;9(11):992-1001.
95. **Park JH, Kim YJ, Binn K, Seo KH**. Spread of multidrug-resistant Escherichia coli harboring integron via swine farm waste water treatment plant. *Ecotoxicology and environmental safety* 2018;149:36-42.
96. **Rehman MU, Zhang H, Huang S, Iqbal MK, Mehmood K et al.** Characteristics of Integrons and Associated Gene Cassettes in Antibiotic-Resistant Escherichia coli Isolated from Free-Ranging Food Animals in China. *Journal of food science* 2017;82(8):1902-1907.
97. **Toleman MA, Walsh TR**. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol Rev* 2011;35(5):912-935.
98. **Partridge SR, Kwong SM, Firth N, Jensen SO**. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clinical microbiology reviews* 2018;31(4).
99. **Harmer CJ, Hall RM, Blokesch M**. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* 2016;1(2).
100. **Venturini C, Hassan KA, Roy Chowdhury P, Paulsen IT, Walker MJ et al.** Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS26-related molecular signature isolated from different Escherichia coli pathotypes from different hosts. *PLoS One* 2013;8(11):e78862.
101. **Canton R, Coque TM**. The CTX-M beta-lactamase pandemic. *Current opinion in microbiology* 2006;9(5):466-475.
102. **Dionisi AM, Lucarelli C, Owczarek S, Luzzi I, Villa L**. Characterization of the plasmid-borne quinolone resistance gene qnrB19 in Salmonella enterica serovar Typhimurium. *Antimicrob Agents Chemother* 2009;53(9):4019-4021.
103. **Doublet B, Praud K, Weill FX, Cloeckaert A**. Association of IS26-composite transposons and complex In4-type integrons generates novel multidrug resistance loci in Salmonella genomic island 1. *J Antimicrob Chemother* 2009;63(2):282-289.
104. **Gilmour MW, Thomson NR, Sanders M, Parkhill J, Taylor DE**. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid* 2004;52(3):182-202.
105. **Mangat CS, Bekal S, Irwin RJ, Mulvey MR**. A Novel Hybrid Plasmid Carrying Multiple Antimicrobial Resistance and Virulence Genes in Salmonella enterica Serovar Dublin. *Antimicrob Agents Chemother* 2017;61(6).
106. **Porse A, Schonning K, Munck C, Sommer MO**. Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Molecular biology and evolution* 2016;33(11):2860-2873.

107. **Iyer A, Barbour, E. , Azhar, E. , Salabi, A. , Hassan, H. , Qadri, I. , Chaudhary, A. , Abuzenadah, A. , Kumosani, T. , Damanhouri, G. , Alawi, M. , Na'was, T. , Nour, A. and Harakeh, S.** Transposable elements in Escherichia coli antimicrobial resistance. . *Advances in Bioscience and Biotechnology* 2013;4:415-423.
108. **Chalmers G, Rozas K, Amachawadi R, Scott H, Norman K et al.** Distribution of the pco Gene Cluster and Associated Genetic Determinants among Swine Escherichia coli from a Controlled Feeding Trial. *Genes* 2018;9(10):504.
109. **Billman-Jacobe H, Liu Y, Haites R, Weaver T, Robinson L et al.** pSTM6-275, a Conjugative IncHI2 Plasmid of Salmonella enterica That Confers Antibiotic and Heavy-Metal Resistance under Changing Physiological Conditions. *Antimicrobial agents and chemotherapy* 2018;62(5).
110. **Heiman KE, Mody RK, Johnson SD, Griffin PM, Gould LH**. Escherichia coli O157 Outbreaks in the United States, 2003-2012. *Emerg Infect Dis* 2015;21(8):1293-1301.
111. **Foxman B**. The epidemiology of urinary tract infection. *Nat Rev Urol* 2010;7(12):653-660.
112. **Johnson JR, Murray AC, Gajewski A, Sullivan M, Snippes P et al.** Isolation and Molecular Characterization of Nalidixic Acid-Resistant Extraintestinal Pathogenic Escherichia coli from Retail Chicken Products. *Antimicrobial agents and chemotherapy* 2003;47(7):2161-2168.
113. **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.
114. **Riley LW**. Pandemic lineages of extraintestinal pathogenic Escherichia coli. *Clin Microbiol Infect* 2014;20(5):380-390.
115. **Cooke EM, Hettiaratchy IG, Buck AC**. Fate of ingested Escherichia coli in normal persons. *J Med Microbiol* 1972;5(3):361-369.
116. **Jakobsen L, Hammerum AM, Frimodt-Moller N**. Virulence of Escherichia coli B2 isolates from meat and animals in a murine model of ascending urinary tract infection (UTI): evidence that UTI is a zoonosis. *J Clin Microbiol* 2010;48(8):2978-2980.
117. **Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K et al.** Escherichia coli ST131-H22 as a Foodborne Uropathogen. *mBio* 2018;9(4).
118. **Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B et al.** Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. *Genome Res* 2015;25(1):119-128.
119. **Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL et al.** Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. *J Infect Dis* 2013;207(6):919-928.
120. **Platell JL, Johnson JR, Cobbold RN, Trott DJ**. Multidrug-resistant extraintestinal pathogenic Escherichia coli of sequence type ST131 in animals and foods. *Vet Microbiol* 2011;153(1-2):99-108.
121. **Mora A, Garcia-Pena FJ, Alonso MP, Pedraza-Diaz S, Ortega-Mora LM et al.** Impact of human-associated Escherichia coli clonal groups in Antarctic pinnipeds: presence of ST73, ST95, ST141 and ST131. *Sci Rep* 2018;8(1):4678.
122. **Mora A, Lopez C, Dabhi G, Blanco M, Blanco JE 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.

123. **Adams-Sapper S, Diep BA, Perdreau-Remington F, Riley LW**. Clonal composition and community clustering of drug-susceptible and -resistant Escherichia coli isolates from bloodstream infections. *Antimicrobial agents and chemotherapy* 2013;57(1):490-497.
124. **Wu XY, Chapman T, Trott DJ, Bettelheim K, Do TN et al.** Comparative analysis of virulence genes, genetic diversity, and phylogeny of commensal and enterotoxigenic Escherichia coli isolates from weaned pigs. *Appl Environ Microbiol* 2007;73(1):83-91.
125. **Schierack P, Walk N, Ewers C, Wilking H, Steinruck H et al.** ExPEC-typical virulence-associated genes correlate with successful colonization by intestinal E. coli in a small piglet group. *Environ Microbiol* 2008;10(7):1742-1751.

Chapter 3: Research Methodology

3.1 Study design

Strain selection

The 68 *E. coli* strains characterised in this thesis are a subset of strains from a larger collection of 171 isolates from 23 piglets. They were collected via rectal swabs from 19-30 day old piglets in 2007 on a farm in NSW, which had a history of neomycin use for treatment of post-weaning diarrhoea. They were isolated on MacConkey agar at the Elizabeth MacArthur Agricultural Institute (EMAI) with up to ten colonies from each piglet being selected. They were originally screened by PCR at EMAI for a suite of ARGs, VAGs and class 1 integrase gene *intI1.* 168/171 (98%) strains were positive for *intI1* suggesting that most were MDR. One hundred *intI1+* strains were initially selected and Illumina whole genome sequenced for study in this thesis and this number was reduced to 68 isolates from 21 pigs after removal of strains that appeared to be clones and poor quality genome sequences.

The *E. coli* strains used in this study originated from a farm where antimicrobials had been used extensively to treat outbreaks of diarrhoeal disease caused by ETEC. This represents a perfect scenario to examine antimicrobial resistance in a setting where selection pressure is high, however it may not reflect the situation on other pig farms with lower antimicrobial use. Whilst efforts were made to isolate multiple colonies from individual pigs, sampling bias is unavoidable and it is impossible to know if the true diversity of a population has been captured by any given sampling methodology. The collection nonetheless exhibited diversity in both phylogenetic background and ARG and VAG carriage so we consider this to be a minor issue. It would have been desirable to sample multiple farms across the country, however this was not feasible. This is somewhat addressed in Chapter 4 where this collection was combined with isolates from another farm that formed part of another student's thesis.

Genomic epidemiology

As discussed in the previous chapter, the emerging discipline of genomic epidemiology is best placed to characterise and understand the dissemination of

antimicrobial resistance and pathogenic lineages of *E. coli* between humans, animals and the environment. It provides far superior resolution, reproducibility, scalability and throughput relative to traditional molecular techniques[1]. Furthermore, most bioinformatics tools used in genomics are open-source, making this field far more accessible than molecular biology and resulting in a large online support community.

However, the field is still in its infancy and there are some limitations worth discussion. Firstly, not all open-source software is peer-reviewed and published so some programs may be more reliable and methodically robust than others. However, most open-source software is developed by experts and reputable university-based institutions, and examination of relevant literature indicates what programs are widely used and accepted. Secondly, genomic analysis typically relies on 'workflows' where multiple programs are required to complete an analysis. Workflows may vary widely depending on the research question being addressed. This results in a lack of standardisation and therefore it is sometimes difficult to compare results between studies. This is particularly an issue with regard to phylogenomics.

Genomic epidemiology requires comparison of bacterial strains from diverse sources, locations and times in order to understand the spread of antimicrobial resistance genes and lineages. The sample size required to make meaningful conclusions far outreaches any individual's ability to sample from the field. The advent of public sequence databases such as GenBank and Enterobase addresses this problem to an extent, though it is not without its pitfalls. For example, approximately one-third of *E. coli* sequences in Enterobase are deposited with metadata for source, year and country of isolation, rendering about two-thirds of the database useless to epidemiological study. Furthermore, a significant bias towards human-origin isolates clearly exists, making it difficult to assess the role of agriculture and environmental isolates. As the number of publicly available sequences increases exponentially and greater consideration is given to non-human sources of bacteria and AMR, these issues should lessen. Nonetheless, despite the small number and diversity of useful sequences to compare, some

striking trends may still be observed and this approach is far more powerful than anything the field of infectious disease epidemiology has had to date.

3.2 Methodology

Full methods and materials can be found in each results chapter. This section will broadly describe and discuss the overall methodology of the thesis.

DNA isolation

DNA isolation for Illumina sequencing was performed using the Bioline™ ISOLATE II Genomic DNA Kit. These kits provide cheap, fast, reproducible DNA isolation suitable for amplification-based Illumina short read sequencing.
Phenol-chloroform extraction was used for Nanopore long read sequencing. Long read sequencing requires pure, concentrated, high molecular weight DNA as there is no amplification step and shearing of DNA yields poor sequence data.

DNA sequencing

Illumina HiSeq

The collection of porcine *E. coli* was sequenced using an Illumina HiSeq 2500. Illumina sequencing is the gold standard of short read sequencing. It provides highly accurate nucleotide base calls however short read lengths are problematic in assembling complex antimicrobial resistance regions that contain a lot of repetitive sequences. This form of sequencing is also relatively inexpensive when compared to newer, long-read technologies. More information on full protocols can be found in the Methods and Materials section of Chapter 4.

Oxford Nanopore Technologies MinION

An ONT MinION was used to perform long read sequencing on two strains described in Chapter 6. Nanopore technology produces reads up to 40kb in length, allowing complex drug resistance regions to be read through and assembled in their entirety. Base calls are not as accurate as Illumina but this problem was resolved by performing hybrid genome assemblies, which use both Nanopore and Illumina reads.

Sequence assembly

Short read assembly

A5-miseq was used to generate *de novo* short read assemblies analysed in Chapter 4[2]. This allowed partial characterisation of resistance regions in the strains, however as with all short read sequencing, complete assembly of resistance regions was not possible.

Hybrid assembly

Unicycler was used to perform hybrid assemblies with Oxford Nanopore reads and Illumina reads. Briefly, Unicycler generates a short read assembly, and then uses the long reads to bridge the gaps, usually yielding closed chromosomes and plasmids. Finally, it performs a polishing step, using the short reads to ensure the base calls of the final contigs are accurate[3]. Bandage is a program that allows visualisation of the quality of these assemblies and was used in conjunction with Unicycler[4]. This method allowed full characterisation of the plasmids present in the sequenced strains (see Chapter 6).

Strain typing and phylogenetic analysis

Multilocus sequence typing (Achtman MLST)

MLST for *E. coli* assigns allele numbers to the nucleotide sequence of seven housekeeping genes and the unique combination of these allele numbers corresponds to a sequence type (ST)[5]. Allele combinations that differ by one or two allele designations belong to the same clonal complex (CC). MLST provides far greater resolution than serotyping, PFGE and other molecular methods and is useful for identifying dominant groups of *E. coli.* However, it is increasingly supplemented by SNP-based whole genome phylogeny approaches for superior delineation of strains. There are numerous viable approaches for inferring phylogenies with whole genome sequences. Whether or not the sequences being analysed are closely related is a major consideration in this respect. Inferring phylogenies typically involves alignment of sequences to a reference genome or gene database, removal of recombinant regions and tree inference using maximum-likelihood Bayesian models of bacterial evolution.

Phylosift

Phylosift, originally designed for the analysis of metagenomic sequences, is optimal for sequences that are diverse as it utilises a concatenated alignment of the amino acid sequence of 37 'elite' gene families, usually found in single copy in bacteria in addition to ribosomal and mitochondrial gene families to infer phylogeny[6]. Phylosift therefore avoids reference genome induced bias and the need for removal of recombinant regions. Phylosift was used for creating a phylogeny of the full collection of porcine *E. coli* in Chapter 4.

SNP-based

SNP-based core genome phylogenies are excellent for reconstructing the evolutionary history of closely related sequences such as strains of the same multilocus sequence type or clonal complex. We used Snippy (https://github.com/tseemann/snippy) for this purpose as it is built on reliable and established tools such as BWA[7] and Freebayes (https://github.com/ekg/freebayes). Furthermore, it is fast and generates full and core alignments of all strains aligned to the same reference genome. It is therefore easy to use with downstream applications such as Gubbins and FastTree2.
We used Gubbins to filter full alignments for recombination. Gubbins identifies regions of recombination based on elevated base substitution densities and is widely used for recombination of haploid genomes[8]. This approach using Snippy and Gubbins was utilised in Chapters 5 and 6.

Inference of phylogenetic trees

We used FastTree2 under a generalised time-reversible model to draw maximum-likelihood trees from all alignments. FastTree2 is fast and accurate for large alignments, such as those we generated in Chapters 4 and 5 in particular[9].

Gene screening

One of the great advantages of WGS in epidemiology is the ability to rapidly identify the presence or absence of any genes of interest. There are two ways to go about this; assembly-based and short read-based.

Assembly-based

Assembly-based gene identification first requires the generation of a de novo assembly followed by identification using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST)[10]. This tool is implemented online and is also available as a command line tool. The latter is particularly useful for screening large collections of genomes against large databases of genes. The advantage of this tool is its speed, particularly when run in parallel on a high-performance computing environment. The only disadvantage is the potential for assembly bias, where misassembly may result in a gene being incorrectly determined present or absent.

Short read-based

Short read-based methods avoid the issue of assembly bias by aligning short reads to reference genes. We used ARIBA for this purpose. Like BLAST, ARIBA can be used with any gene database and can additionally identify novel alleles and amino acid changes. It also produces summary output with simple gene presence or absence as an advantage over the sometimes convoluted output generated by BLAST[11].

Gene Databases

The Centre for Genomic Epidemiology (CGE) maintains a number of excellent and extensive databases of genes relevant to genomic epidemiology. These are available as web-based screening tools, however they are also compatible with ARIBA. Databases used in this thesis include ResFinder, VirulenceFinder, PlasmidFinder, pMLST and SerotypeFinder[12-15]. We also utilised a custom database of genes derived from relevant literature to supplement the CGE databases.

Genome annotation

In order to fully characterise plasmids examined in Chapter 6, we utilised Rapid Annotation using Subsystems Technology toolkit (RASTtk) via the Pathosystems Resource Integration Centre's (PATRIC) online Bacterial Bioinformatics Resource Centre (https://patricbrc.org/)[16]. This allowed identification of hypothetical and

predicted proteins in addition to ARGs, VAGs and plasmid related genes we identified with BLAST.

Comparative analysis

SnapGene

SnapGene (GSL Biotech) is a proprietary desktop program for visualisation and annotation of nucleotide and amino acid sequences in a variety of formats. It is useful for manual editing of automated annotation and generates quality maps of plasmids for export.

BLAST Ring Image Generator (BRIG)

BRIG is a Java-based tool for BLAST-based comparison of nucleotide sequences. It is useful for visualising sequence homology and deletions in chromosomes and plasmids[17]. One shortcoming is its inability to visualise insertions or additional gene content in sequences selected for comparison relative to the reference sequence.

Data visualisation

Visualising phylogenetic trees in conjunction with relevant metadata and gene carriage information is critical in genomic epidemiology. For phylogenetic trees and metadata we used the Interactive Tree of Life (iTOL) web server (https://itol.embl.de), which allows tree extensive manipulation and programmatic editing of metadata. To visualise gene screening results we used the R package ggtree, which also allows extensive tree manipulation and generation of heatmaps[18]. Whilst it is more flexible than iTOL, it is less user-friendly.

3.3 References

1. **Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM et al.** Genomic Microbial Epidemiology Is Needed to Comprehend the Global Problem of Antibiotic Resistance and to Improve Pathogen Diagnosis. *Front Microbiol* 2016;7:843.
2. **Coil D, Jospin G, Darling AE**. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015;31(4):587-589.
3. **Wick RR, Judd LM, Gorrie CL, Holt KE**. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13(6):e1005595.
4. **Wick RR, Schultz MB, Zobel J, Holt KE**. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015;31(20):3350-3352.
5. **Wirth T, Falush D, Lan R, Colles F, Mensa P et al.** Sex and virulence in Escherichia coli: an evolutionary perspective. *Molecular Microbiology* 2006;60(5):1136-1151.
6. **Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM et al.** PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014;2:e243.
7. **Li H, Durbin R**. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009;25(14):1754-1760.
8. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA et al.** Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research* 2015;43(3):e15.
9. **Price MN, Dehal PS, Arkin AP**. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5(3):e9490.
10. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. Basic local alignment search tool. *Journal of molecular biology* 1990;215(3):403-410.
11. **Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J et al.** ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3(10):e000131.
12. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
13. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS et al.** Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol* 2014;52(5):1501-1510.
14. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F**. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2015;53(8):2410-2426.
15. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al.** Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67(11):2640-2644.
16. **Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S et al.** RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015;5:8365.
17. **Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA**. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, journal article 2011;12(1):402.

18. **Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y et al.** ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;8(1):28-36.

Chapter 4: Porcine commensal *Escherichia coli*: a reservoir for class 1 integrons associate with IS*26*

4.1 Declaration

Cameron J. Reid[1#], Ethan R. Wyrsch[1#], Piklu Roy Chowdhury[1,2], Tiziana Zingali[1], Michael Y. Liu[1], Aaron E Darling[1], Toni A. Chapman[2], and Steven P. Djordjevic[1*]

[1]The i3 institute, University of Technology Sydney, Ultimo, NSW 2007, Australia

[2]NSW Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Private Bag 4008 Narellan, NSW, 2567, Australia.
*corresponding author
# Authors contributed equally

Published in Microbial Genomics, December 2017, doi: 10.1099/mgen.0.000143

Author contribution statement

The primary authorship of this publication was shared by C.J.R. and E.R.W. as the paper contained two separate collections of strains (F2 and F1 respectively) that formed the basis of their theses. C.J.R. generated gene screening data for both F1 and F2. E.R.W. generated the phylogeny for F1 and F2. Analysis of all data was performed together.

M.Y.L. prepared samples for sequencing. Assembly was performed by A.E.D. T.Z. performed phenotypic CDS testing. The study was conceived by S.P.D., P.R.C., and T.A.C. T.A.C. was responsible for provision and preliminary PCR characterisation of the strains. The manuscript was written and edited by C.J.R., E.R.W., and S.P.D.

4.2 Porcine commensal *Escherichia coli*: a reservoir for class 1 integrons associated with IS*26*

This publication addresses Aims 1 and 2 of the thesis. Please note, figure numbers have been edited from the publication so as to match the layout of the thesis, *i.e.* Fig. 1 is now Fig. 4.1 *etc.* Supplementary data has been renumbered in the same manner. Please view Figures as PDFs attached to this thesis.

4.2.1 Abstract

Porcine faecal waste is a serious environmental pollutant. Carriage of antimicrobial-resistance genes (ARGs) and virulence- associated genes (VAGs), and the zoonotic potential of commensal Escherichia coli from swine are largely unknown. Furthermore, little is known about the role of commensal *E. coli* as contributors to the mobilization of ARGs between food animals and the environment. Here, we report whole-genome sequence analysis of 103 class 1 integron-positive *E. coli* from the faeces of healthy pigs from two commercial production facilities in New South Wales, Australia. Most strains belonged to phylogroups A and B1, and carried VAGs linked with extraintestinal infection in humans. The 103 strains belonged to 37 multilocus sequence types and clonal complex 10 featured prominently. Seventeen ARGs were detected and 97 % (100/103) of strains carried three or more ARGs. Heavy-metal-resistance genes *merA*, *cusA* and *terA* were also common. IS*26* was observed in 98 % (101/103) of strains and was often physically associated with structurally diverse class 1 integrons that carried unique genetic features, which may be tracked. This study provides, to our knowledge, the first detailed genomic analysis and point of reference for commensal *E. coli* of porcine origin in Australia, facilitating tracking of specific lineages and the mobile resistance genes they carry.

4.2.2 Data summary

One hundred and forty-three whole-genome sequences of porcine faecal Escherichia coli sequenced in this project have been deposited at the European Molecular Biology Laboratory (EMBL) European Nucleotide Archive under study accession number PRJEB21464. For individual sample accession numbers, please refer to Table S4.1 (available in the online version of this article). Further strain data is available in Tables S4.2–S4.6.

4.2.3 Introduction

Escherichia coli is the most frequently isolated Gram-negative pathogen affecting human health [1]. They are frequently resistant to multiple antibiotics and modeling studies forecast that multidrug resistant (MDR; resistant to three or more classes of antimicrobials) *E. coli* infections will account for 30% of 10 million fatal MDR infections annually by 2050 [2]. In addition to the pathogenic variants,

commensal *E. coli* comprise an important component of the gut microbiota. *E. coli* are shed into the environment in high numbers. For example, each gram of faeces from commercially reared pigs contains between 10^4 – 10^8 *E. coli* [3]. It is important to understand the characteristics of these *E. coli* given the huge quantities of faeces generated and disseminated by intensive pig production. China, the world's largest producer of swine, produces an estimated 0.618 billion to 1.29 billion metric tonnes of faeces each year [4, 5].

Pathogenic *E. coli* are broadly divided into intestinal pathogenic *E. coli* (IPEC) and extraintestinal pathogenic *E. coli* (ExPEC). ExPEC have a faecal origin, having persisted asymptomatically in the gut before opportunistically colonising extraintestinal sites where they cause a diverse range of diseases including urinary tract infections, pyelonephritis, wound infections, sepsis and meningitis [6]. ExPEC are thought to have foodborne reservoirs and may enter the food chain via a number of sources [7-11]. The zoonotic potential of commensal porcine *E. coli* as a source of ExPEC that cause disease in humans is unknown. ExPEC cannot be reliably detected in a diagnostic test as they are yet to be shown to possess unique identifying features relative to other pathotypes of *E. coli* [12]. Instead, as we aim to do here, whole genome sequencing can be used to discriminate strains indistinguishable by other methods and identify any genetic relationships between *E. coli* strains isolated from pigs and humans.

Horizontal gene transfer, mediated by mobile genetic elements, plays an important role in the evolution of *E. coli*. Commensal or pathogenic bacteria may, in a single horizontal gene transfer event, acquire a mobile genetic element carrying multiple antimicrobial resistance genes (ARGs), virulence-associated genes (VAGs), and other genetic cargo that encode traits that offer a niche advantage [13-17]. The release of MDR commensal *E. coli* into the environment, such as when pig faeces are used as manure, facilitates horizontal transfer of resistance and virulence genes into other microbial communities in a manner that is poorly understood. ARGs cluster on mobile genetic elements and form complex resistance regions that are often independently mobile. Indirect selection pressure can, in the absence of antibiotic use, lead to the persistence of transferred genes. For example, heavy

metals such as copper and zinc in feed formulations for food animals select for ARGs that co-localise with metal resistance genes [18, 19]. Selection pressure afforded by any one of a number of antibiotics and heavy metals (zinc, cadmium, mercury) that contaminate faecal waste or those used in food-producing and hospital environments is sufficient to select for the retention and spread of complex resistance regions [20, 21]. Understanding of how ARGs assemble on mobile genetic elements and the extent to which these then traffic through human, food animal and environmental reservoirs remains limited.

Class 1 integrons are a reliable proxy for the presence of multiple ARGs within bacteria in clinical and veterinary settings [22]. They are gene capture and expression elements that can integrate AMR gene cassettes from the environmental resistome and express them via a promoter residing in the class 1 integrase gene. They are often mobilised by mercury resistance transposons belonging to the Tn*21* family, which have been disseminated globally on a wide variety of conjugative plasmid backbones [23]. Resistance genes can also be acquired, lost and rearranged in bacteria by genetic events that involve insertion sequences (IS) such as IS*26*, IS*Ecp1*and IS*CR1* [24-27]. IS*26* is prominent in this regard due to its unique mechanisms of transposition (conservative and replicative), ability to recognise itself, lack of copy number control and ability to mobilise a wide range of ARGs [15, 26, 28-30]. Furthermore, IS*26* is recognised to play a key role in: i) the evolution of plasmids and genomic islands that carry combinations of VAGs and ARGs [16, 17, 31, 32]; ii) driving the formation of cointegrate plasmids encoding VAGs and ARGs [33] and; iii) initiating deletions in large multidrug resistance plasmids that enhance plasmid stability and expand host range [34].

Infectious disease management relies on surveillance of antimicrobial resistance and emerging pathogens using a One Health approach. There is currently no published data available that records whole genome sequence-based phylogeny, ARG or VAG carriage in commensal *E. coli* from Australian pigs and only one comparable study is available from overseas [35]. Here, for the first time, we present whole genome sequence analysis of 103 class 1 integron-positive

commensal *E. coli* from pigs commercially-reared in Australia. We present data characterising their phylogenetic diversity, carriage of VAGs, ARGs and an analysis of the class 1 integrons they carry.

4.2.4 Methods

Management of farms and animals

The study was conducted using *E. coli* sourced separately from two pig production farm systems located approximately 250 km apart. Farms were designated descriptors F1 and F2. Isolate numbers consist of farm number, a pig number and a letter designating a single isolate from that pig (i.e. F1_404D indicates farm 1, isolate D from pig 404). At both farms, pigs are intensively housed and kept in total confinement. Both farms have used neomycin in the past for the treatment of diarrhoeal disease. No antibiotics were being used during the first sampling time at F1, however the pigs sampled at the second sampling time had received a course of neomycin (see below). No antibiotics had been administered to the pigs at F2 prior to sampling.

Escherichia coli strains used in the study

Escherichia coli isolates were collected via rectal swab sampling of pigs between 19 and 30 days of age. At farm 1, rectal swabs were collected in May 2007 from pigs during an outbreak of diarrhoeal disease but prior to treatment with neomycin. These pigs were subsequently removed from the shed. The causative agent of the outbreak was unknown. A new batch of healthy sows and their piglets were transferred to this shed and the sows were given neomycin in-feed. Once the piglets were weaned they also received neomycin in-feed for 7-10 days. The second sampling occurred on these piglets in June 2007 after the course of antibiotics. At farm 2, rectal swabs were performed on healthy weaners that were not treated with antibiotics.

E. coli were isolated at the Elizabeth MacArthur Agricultural Institute (EMAI). Up to ten *E. coli* colonies were selected from individual pigs using MacConkey Agar. The total collection from farm 1 was 164 isolates from 33 pigs whilst farm 2 was 171 isolates from 23 pigs. All strains were screened by PCR for the class 1

integrase gene *intI1*. This screening indicated that 117/164 (71%) *E. coli* from farm 1 and 168/171 (98%) from farm 2 carried *intI1.* Initially, 50 *intI1* positive isolates from F1 and 100 *intI1* positive isolates from F2 were selected for whole genome sequencing. Two enterotoxigenic *E. coli* strains, M10 and ETEC286/3, which were submitted to EMAI from Australian veterinary services, as clinical, pig-derived strains were also sequenced and included in the phylogenetic analysis as reference strains.

Storage

All strains were freshly cultured in LB broth and frozen as glycerol stocks made using 500 µl of M9 salts solution and 500 µl of 50% glycerol and stored at -80°C. All strains were cultured in LB broth prior to isolation of gDNA used for sequencing.

DNA extraction, whole genome sequencing and assembly

Genomic DNA was extracted using the ISOLATE II Genomic DNA Kit (Bioline, Eveleigh, Australia) following the manufacturers standard protocol for bacterial cells and stored at -20°C. Whole genome sequencing libraries were prepared from separate aliquots of sample gDNA using the Illumina Nextera DNA kit with modifications. In brief, the gDNA was first quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Scoresby, Australia). All sample gDNA concentrations were standardised to equal concentration to achieve uniform reaction efficiency in the tagmentation step. Standard Illumina Nextera adaptors were used for sample tagmentation. The PCR-mediated adapter addition and library amplification was carried out using customized indexed i5 and i7 adaptor primers (IDT, Coralville, IA, USA), which were developed based on the standard Nextera XT Indexed i5 and i7 adapters (e.g. N701-N729 and S502-S522). Libraries were then pooled and size selected using SPRI-Select magnetic beads (Beckman Coulter, Lane Cove West, Australia). Finally, the pooled library was quality checked and quantified on an Agilent Bioanalyzer 2100 using the DNA HS kit (Agilent, Santa Clara, CA, USA). Whole genome sequencing for the majority of F1 strains and ETEC strains was performed as previously reported [36] using an Illumina MiSeq sequencer and MiSeq V3 chemistry. Whole genome sequencing of the remaining F1

and F2 strains was performed using an Illumina HiSeq 2500 v4 sequencer in rapid PE150 mode (Illumina, San Diego, CA, USA). Sequence read quality was initially assessed using FastQC version 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Illumina raw reads passing quality control were assembled into draft genome sequences using the A5 assembly pipeline version A5-miseq 20140604 [37]. Genome sequences have been deposited in the European Nucleotide Archive with study accession number PRJEB21464. Accession numbers for each sample are listed in Table S4.1.

Strain selection

Sequence data was successfully generated for 141 strains and these were screened by BLAST for *intI1,* ARGs and VAGs and subjected to Phylosift analysis as described below. These analyses indicated 12 strains were negative for *intI1* and that a number of clones were isolated from individual pigs. We therefore excluded *intI1*-negative strains and selected representatives of the clonal isolates, thereby excluding a further 26 strains. The subset of strains that were sequenced were identified as F1 + F2 (n = 103 from 42 pigs). This subset consisted of 35 strains from 21 pigs sampled at farm 1 and 68 strains from 21 pigs sampled at farm 2; among the F1 strains, 17 were disease-associated from 12 pigs (isolate numbers 1-30, designated 'Disease' in Table S4.1-S4.5) and 18 were isolated from 11 healthy pigs (isolate numbers 365-409, designated 'Healthy' in Table S4.1-S4.5). Only 11 isolates in the collection carried toxin genes (*eltA;* n=2*, eltB;* n=2*, stA;* n=0*, stB;* n=11) associated with porcine enterotoxigenic *E. coli* (ETEC) and no ETEC adhesins were detected*.* Notably, only 5 of these were from diseased pigs whilst 6 were from healthy pigs. This highlights the role that host factors such as stress and immune health play in the manifestation of pre- and post-weaning diarrhoea in pigs and we therefore argue that this collection should be considered commensal.

Assembly Statistics

Comprehensive assembly statistics for 143 sequenced porcine-derived *E. coli,* (141 + 2 ETEC) are available in Table S4.1. Isolates not included in this study are highlighted grey. The number of scaffolds per genome ranged from 29 to 1571,

with a mean of 235. Each genome sequence had a median sequencing coverage of at least 20 ×, with a maximum of 94 × and mean of 54 ×.

Phenotypic resistance testing

F1 strains were tested at EMAI using the calibrated dichotomous susceptibility test (CDS) for resistance to 12 antibiotics [38]. The following were tested: ampicillin (25 µg), cefoxitin (30 µg), nalidixic acid (30 µg), ciprofloxacin (2.5 µg), imipenem (10 µg), sulphafurazole (300 µg), trimethoprim (5 µg), tetracycline (10 µg), neomycin (30 µg), gentamicin (10 µg), azithromycin (15 µg) and chloramphenicol (30 µg). F2 strains were tested for resistance to antibiotics at the ithree institute, University of Technology Sydney using the same method and panel of antibiotics as the F1 collection. F2 strains were also tested with streptomycin (25 µg) and kanamycin (50 µg) (Table S4.2).

Gene identification and serotyping

Resistance, virulence and plasmid-associated genes were identified using local BLASTn v2.2.30+ [39] searches with an e-value of 1.0×10^{-3} (Table S4.3-S4.5). Gene databases used were ResFinder, PlasmidFinder, ISFinder, SerotypeFinder and VirulenceFinder [Data references 1-5] [40-44]. Our virulence database was supplemented with additional virulence genes from GenBank available in Table S4.6. Genes were considered present if the subject nucleotide sequence was > 90% identical over 100% of the length of the query sequence. BLAST hits with > 90% identity but covering less than 100% of the query were considered positive if they were truncated by a scaffold break or insertion. Integrons were characterised in SnapGene (GSL Biotech, USA) using BLASTn output. The collection was then retroactively screened for characterised integrons using BLASTn. Where strains carry two *intI1* genes, de novo assembly software is unable to assemble the two complete integrons with Illumina short read data, as it cannot determine which cassette array belongs to which *intI1* copy. The presence of two integrons in strains in this collection was therefore initially inferred by BLAST identification of their cassette arrays and downstream regions (e.g IS*26* deletion signatures) and then confirmed by read-mapping using Bowtie2 and Tablet [45, 46].

Phylogrouping and eMLST

E. coli phylogroups were determined using the scheme published by Clermont *et al,.* [47]. The genes *chuA* (gb|U67920.1), *yjaA* (gb|NC_000913.3) and the DNA fragment TspE4.C2 (gb|AF222188.1) were sourced from GenBank and identified *in silico* using BLASTn. Multilocus sequence typing (MLST) was performed *in silico* using the PubMLST database (http://pubmlst.org/) and the Achtman *E. coli* MLST scheme (http://mlst.warwick.ac.uk/mlst/).

Phylogenetic Analyses

Maximum-likelihood phylogenetic distances between genomes were analysed using the PhyloSift pipeline [48], and a tree was generated using FastTree2 [49]. The tree was visualised using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and iTOL (https://itol.embl.de/). The FastTree2 protocol was modified to resolve short branches as described previously [50].

4.2.5 Results

Our study collection consisted of 103/335 (31%) strains of *E. coli* isolated from rectal swabs of pigs from two farms in New South Wales, Australia that were PCR-positive for the class 1 integron integrase gene, *intI1.* Initial screening indicated that 117/164 (71%) *E. coli* from farm 1 and 168/171 (98%) from farm 2 carried *intI1*.

Population structure of E. coli isolated from porcine rectal swabs

Strains in our study collection were classified by phylogrouping, *in silico* MLST, and *in silico* serotyping. The majority of the strains in our study collection 74/103 (72%) belonged to phylogroup A while the remainder belonged to phylogroup B1 (18; 17%), phylogroup B2 (5; 5%), and phylogroup D (6; 6%).

We identified 37 distinct sequence types, 21 of which were previously isolated from swine, as reported by the *E. coli* MLST database (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli; accessed June 2017). Only seven sequence types were common to both F1 and F2. The most prominent sequence types were ST10, ST361, ST641, ST542, ST48 and ST218. Twenty-five STs were

represented by a single isolate. Six strains with a single SNP in a reference allele were assigned putative sequence types (denoted by an asterisk in Fig. 4.1 and Table S4.3). A designation of non-typable (NT) was given to the 5 remaining strains for which one or more alleles could not be determined.

In silico O:H typing using SerotypeFinder predicted 47 serotypes for 85 strains. The remaining 18 strains were O-non-typable with 10 different H types (Fig. 4.1 and Table S4.3). In general, strains of any given ST carried the same O:H alleles, though intra-sequence type variability was observed among eight sequence types (ST10, ST48, ST218, ST542, ST641, ST302, ST4630 and ST1437)

Phylogenetic analysis

To determine genetic relatedness, we used PhyloSift, FastTree2 and FigTree v1.4.2 to generate and visualise a mid-point rooted, maximum-likelihood phylogenetic tree containing the F1 + F2 pig *E. coli* draft whole genome sequences, two ETEC strains (ETEC286_3 and ETECM_10) and 4 pig-pathogenic *E. coli* complete genome sequences: *E. coli* UMNK88 (NC_017641.1), UMNF18 (NZ_AGTD01000001.1), PCN033 (NZ_CP006632.1) and PCN061 (NZ_CP006636.1) (Fig. 4.1). Tree topology was highly congruent with Achtman MLST and *in silico* serotyping, grouping strains by sequence type, and then further by serotype. Clade structure was generally congruent with phylogroup analyses; however, seven strains belonging to phylogroups B2 and D formed a separate clade. We identified three major clades, with the seven B2/D phylogroup strains forming Clade 1. Clade 2 consisted almost exclusively of phylogroup B1 strains, ST641 was the dominant sequence type however one phylogroup A strain (F1_4A) was an unexpected member of this clade. Clade 3 was composed of two separate sub-clades, one consisted of six B2 and D strains (three ST302, two ST1508 and a non-typable) and the other exclusively containing phylogroup A strains (ST10 and sequence types within CC10, as well as ST361 and ST542, strains that were common in our study collection).

Antimicrobial resistance genes and heavy metal resistance genes

We identified a total of 17 ARGs in the collection and strains carried between 1 and 15 ARGs each. 100/103 (97%) strains carried 3 or more resistance genes. Surprisingly, strains belonging to phylogroup A carried the highest average number of ARGs (10 per strain). Strains belonging to phylogroup B1, B2 and D each carried an average of 8 ARGs per strain, respectively. The most common ARGs among the strains in our collection were the penicillin resistance gene, *bla*$_{\text{TEM-1,}}$ (84; 82%); *aphA1,* encoding resistance to kanamycin and neomycin (76; 74%), the co-linked streptomycin resistance genes, *strA* and *strB* (73; 71%), and the tetracycline resistance gene *tetA* (73; 71%). Quinolone resistance genes *oqxAB,* which typically localize on plasmids (27; 26%) were less frequently identified. Genes encoding extended spectrum β-lactamases (ESBL), extended spectrum carbapenemases (ESC) and resistance to macrolides were not detected. Heavy metal resistance genes including the copper resistance gene *cusA* (103; 100%), the Tn*21* mercury resistance gene *merA* (71; 69%) and the tellurite resistance gene *terA* (40; 39%) were identified frequently (Fig. 4.2).

Five ARGs were identified as gene cassettes carried by class 1 integrons (Fig. 4.3). Cassettes carried by the majority of strains included those conferring aminoglycoside resistance, *aadA1* (69; 67%) and *aadA2* (72; 70%); chloramphenicol resistance, *cmlA,* (60; 58%) and trimethoprim resistance, *dfrA12* (62; 60%) and *dfrA5* (51; 50%). Among sulphonamide resistance genes, *sul3* was identified in more strains (62; 60%) than *sul1* (48; 47%) or *sul2* (46; 45%) (Fig. 4.2 and Table S4.3). *sul1* and *sul3* were associated with integrons (Fig. 4.3).

Multidrug resistant porcine E. coli carry structurally-diverse class 1 integrons

Among our study collection, we sought to characterise the diversity of class 1 integrons present. It is challenging to assemble complete sequences for such regions using Illumina sequence data because of the presence of repeated elements. However, we identified numerous structurally diverse class 1 integrons, hereafter referred to as integrons (a-j) (Fig. 4.1 and Fig. 4.3). Notably IS*26* altered the 3′ region in six of the most common structures (d-i).

Four different class 1 integrons, (g-j) carried a *sul3* gene. The first time *sul3* was linked with *E. coli* from a food-animal source in Australia was in 2015 in a highly virulent porcine ST4245 ExPEC strain [50]. Moreover, *sul3* was first reported in a human in Australia in 2017 in a commensal *E. coli* ST95 [51]. In (g) and (h), the *sul3* module, which comprises a putative transposase *tnp440, sul3,* two hypothetical proteins (*orfA* and *orfB*), and 260 bp of the macrolide efflux gene *mefB* truncated by IS*26*, was the same. Integrons (g) and (h) differed from each other in their respective cassette arrays. Integron (i) differed from (g) and (h) both in its *sul3* module, which carried an additional copy of IS*26,* length of the *mefB* gene fragment (111 bp) and an insertion of an IS*1203*-like element in *qacH*. In (j), an IS*26* insertion leaves only 197bp of *intI1* remaining, *mefB* is absent and an IS*1*-like element is adjacent to *orfB.* Only three of the integrons (a-c) among our strain collection carried a *sul1* gene. Screening indicated that at least 22 strains carry two integron structures. The most common co-carriage pattern was (d, i) (14/22) though (b, i) (2/22), (d, j) (4/22) and (d, g) (2/22) also occurred (Fig. 4.1 and Table S4.3). Eight sequence types carried more than one integron including predominant types ST10, ST361 and ST542 (Table S4.3).

Virulence-associated genes in porcine faecal E. coli

To assess the virulence potential of commensal pig *E. coli* strains in our collection, we screened for a total of 94 genes that have been associated with either intestinal disease or extraintestinal disease caused by *E. coli* pathotypes. Twenty-nine of these genes were present in at least one strain (Fig. 4.2 and Table S4.4). All strains possessed between 3 and 16 VAGs. The average number of VAGs for each phylogroup was A: 5, B1: 9, B2: 11, and D: 9. The VAGs were present in diverse gene combinations between and within sequence types. Most VAGs were typical of extraintestinal *E. coli* pathotypes (ExPEC), whilst ETEC toxin gene (*eltA, eltB, stA, stB*) carriage was only observed in 11 strains and no ETEC adhesins were present.

Plasmid incompatibility groups of porcine faecal E. coli

We screened the collection for plasmid replication-associated genes from nine plasmid incompatibility groups that are commonly associated with carriage and mobility of ARGs. IncF was the most common replicon (89; 72%) followed by IncX

(61; 59%) and IncHI2 (43; 42%). All replicons were present across multiple sequence types (Fig. 4.2 and Table S4.5).

4.2.6 Discussion

Globally, there is a poor representation of genomic sequences for commensal *E. coli* isolated from the faeces of pigs, and none in Australia. Here, for the first time, we sequenced the genomes of *E. coli* isolated from the faeces of predominantly healthy pigs and determined their Clermont phylogroup, MLST (Achtman), serotype as well as carriage of ARGs and VAGs. The phylogenetic relationships shared by the 103 strains, the types of resistance genes that reside within the class 1 integrons, and the structures of class 1 integrons were also investigated. Despite sampling only two commercial piggeries, we identified a wide variety of multilocus sequence types. The diversity of isolates differed to previous studies on *E. coli* in pigs [35, 52] and this may be due to our selection of *intI1*-positive strains or simply reflect geographical differences. Our findings suggest that commensal *E. coli* populations residing within the faeces of pigs are often resistant to multiple antimicrobial agents and carry numerous VAGs. Notably, we also identified genetic epidemiological markers for tracking antimicrobial resistance loci residing on mobile genetic elements in commensal *E. coli*.

Commensal E. coli lineages are associated with disease

The dominant lineages in our collection were phylogroup A *E. coli* belonging to sequence types residing within CC10, particularly ST10, ST48, ST218. ST10 has previously been reported as the dominant sequence type from pigs in Germany, Denmark, Ireland and Spain [3, 35, 52-54]. Our data and the observations of others suggest *E. coli* of CC10 sequence type may be opportunistic, MDR pathogens with a broad animal host range. *E. coli* CC10 can colonise humans, swine, poultry, dogs, migratory birds, rodents, camels and cattle [9, 12, 55-61]. *E. coli* CC10 can also be isolated from raw and treated wastewater, and from urban streams [8]. *E. coli* CC10 is increasingly associated with intestinal disease in humans [62, 63] and extraintestinal infections in pigs [64, 65], dogs [57], and humans, including UTI, pyelonephritis and sepsis [9, 66-68]. *E. coli* CC10 are often MDR, and the resistance genes they carry can encode resistance to extended-spectrum beta-lactams [69,

70]. ST10 is a noted ExPEC sequence type in humans and has been identified in food animals, retail meats, and the environment [58, 71-74]. The core attributes of ST10 that enable it to colonise diverse niches remain unknown. The phylogenetic diversity we observed within porcine faecal ST10 suggests that such attributes may vary between strains. Whole genome sequence analysis of *E. coli* ST10 genomes from different regions of the world and from different hosts is needed to understand the full diversity and success of this sequence type.

MDR porcine E. coli carry structurally-diverse class 1 integrons

Notably, *sul3* was the most frequently identified *sul* gene in our collection and three different *sul3*-containing integron structures were identified. Carriage of class 1 integrons possessing *sul3* has been observed in disease-associated and commensal *E. coli* isolates from animals and humans, as well as in bacterial species other than *E. coli* from different countries [75-77]. In Australia, the carriage of *sul3* by *E. coli* has been reported infrequently, although it has been identified in several uropathogenic *E. coli* isolates [78], in a highly virulent porcine ST4245 ExPEC strain [50], and in a human commensal ST95 *E. coli* on a virulence plasmid that carries multiple ARGs and VAGs [14]. In Europe, class 1 integrons containing *sul3* have been observed in commensal *E. coli* from both humans and animals, indicating they are widely disseminated in a variety of *E. coli* lineages [14, 79-82]. Structures similar to ours have also been reported in different *Salmonella enterica* serovars, suggesting inter-species transfer of class 1 integrons carrying *sul3* may have occurred [75].

The potential role for *sul3*-integrons in intra- and interspecies exchange of antibiotic resistance makes it desirable both to understand their evolution and to track their movement through bacterial populations. In Fig. 4.4, we provided a model that could explain the micro-evolutionary events that created the novel *sul3*-integron depicted in structure (i). This integron likely evolved from a progenitor similar to one described by Curiao *et al* in a human-derived ESBL-positive *E. coli* on an IncI1 plasmid from Spain (gb|HQ875016.1), as this is the only report to describe IS*26* adjacent to *sul3* [76]. Conceivably, the novel structure (i) emerged from insertion of a second copy of IS*26,* which further truncated *mefB,*

followed by an inversion event. To our knowledge, this is the first study to identify a 111-bp *mefB* variant. Integron (i) was observed within the collection in 26 *E. coli* strains of different sequence types, suggesting horizontal transfer of a mobile element(s) carrying the integron though we were unable to determine which mobile elements are responsible for this. Further work is needed to examine this hypothesis.

IS*26*-mediated deletions of the *mefB* can be used to track *sul3*-containing integrons and additional resistance genes they may acquire due to the unique ability of IS*26* to target itself [26]. A number of different truncated variants of the *mefB* gene are carried by *sul3*-integrons found in human and animal derived *E. coli* [14, 75-77]. Our data suggests the class 1 integrase upstream of the *sul3* module is likely to be functional based on the presence of different antibiotic cassette arrays associated with a 260-bp *mefB* deletion (g, h). BLASTn analysis identified *sul3* integrons carrying Δ*mefB* with an identical 260-bp deletion in porcine isolates P328.10.99.C2 (gb| FJ196386.1) and P528.10.99.C4 (gb| FJ196388.1) from Great Britain, though the associated cassette arrays were not completely characterised [77]. Furthermore, plasmid pCAZ590 (gb| LT669764.1) isolated from poultry in Germany carried an identical integron (*estX-psp-aadA2-cmlA-aadA1-qacI-tnp440-sul3-orf1-orf2-ΔmefB:*260bp-IS*26*) to 4(h) with an additional $bla_{\text{SHV-12}}$ gene 73 bp upstream of IS*26* [83]. Although the evolutionary events that lead to this derivative structure are not known, this plasmid illustrates how IS*26* augmented integrons continue to evolve and acquire genes that confer resistance to critically important human antibiotics.

The deletion event in the 3′-*CS* of the integron depicted in (d) (*dfrA5*-IS*26*) may serve as another genetic signature for tracking resistance genes, and bacteria that carry them, through different hosts and environments [15, 84-86]. Previously we observed the integron structure (d) on plasmids carrying VAGs in atypical EPEC strains isolated from cattle with gastrointestinal disease and *E. coli* strains linked to EHEC O26:H- isolated from a human patient with haemorrhagic colitis [16, 17]. In each of these earlier cases, the IS*26* that interrupted the 3′-*CS* of the integron formed part of the left boundary of Tn*6026,* an IS*26*-flanked, globally disseminated

transposon that harbours multiple ARGs [15-17, 87, 88]. Twenty-seven strains carrying integron (d) possess the resistance genes present in Tn*6026* (*bla*$_{\text{TEM}}$, *sul2, strAB, aphA1*) suggesting this transposon is also carried in our collection though further studies are necessary to confirm this. This again highlights that tracking IS*26* deletions is useful not only for tracking the integrons they interrupt, but also additional resistance genes that may be acquired in association with the IS*26.*

The carriage of more than one integron in a number of prominent sequence types in the collection suggests that plasmid or transposon-mediated horizontal transfer of resistance determinants may occur within the microbiota of the porcine gut. This transfer is likely mediated by plasmids present in the collection, though transposons and IS elements may be involved. Long-read sequencing is required to test this hypothesis.

Zoonotic potential of commensal. E. coli from swine

In considering the zoonotic potential of pig faecal *E. coli*, we determined the proportion of strains in our collection that carried IPEC and ExPEC VAGs. A limitation of investigating zoonotic potential for extraintestinal disease is the genetic redundancy identified in the virulence attributes from ExPEC. A recent study suggested that the number of virulence factors carried by an ExPEC strain is the only independent factor that can explain extraintestinal virulence in a mouse model of sepsis [89]. Our collection contained two strains possessing large numbers of VAGs, belonging to ST131 and ST117, representative of pandemic ExPEC clones that cause hospital and community-acquired infections in humans worldwide [58, 90, 91]. They have both been linked with poultry and have only rarely been isolated from porcine sources [9, 58]. The single ST131 strain in our porcine collection carried 10 ARGs and 16 VAGs. The ST117 strain carried eight ARGs and 16 VAGs, including the full array of iron acquisition genes *fyuA, irp2, ireA, iroN, iutA, iucD* and *sitD.* Several of these genes are typically encoded on virulence plasmids circulating in APEC [92] and this profile is similar to ST117-O111:H4 strains from poultry previously reported by Mora *et al,* [93]. The presence of ST117 and ST131 in our collection is intriguing and warrants further investigation.

Most of the VAGs identified in our collection were those associated with the ability to cause extraintestinal disease in humans as well as intestinal persistence [6, 94]. Carriage of genes that are under positive selection in uropathogenic *E. coli* (UPEC) [95], such as heat-stable agglutinin gene *hra* [96], murine uroepithelial cell adhesin gene *iha* [97], iron acquisition genes, *fyuA, iutA, iucD* and *sitD;* and the serum survival genes *iss* and *traT* suggest that some strains may be capable of causing extraintestinal disease in humans. Conversely, it also highlights how many ExPEC VAGs can be considered important intestinal fitness factors. The most intriguing IPEC VAG was intimin gene *eaeA*, found in 8 strains, that is characteristic of several intestinal *E. coli* pathotypes, including EHEC, EPEC and atypical EPEC [98]. These strains also carried ExPEC VAGs and may represent hybrid pathotypes.

The frequency of VAGs in phylogroups A and B1, an average of 5.4 and 8.9 VAGs per isolate respectively, was unexpected because *E. coli* belonging to phylogroups A and B1 are considered to have low virulence potential [99, 100]. The carriage of multiple VAGs in pig *E. coli* is consistent with earlier studies [101, 102]. In China, ExPEC have been isolated from a variety of tissues and bodily fluids of pigs with septicaemia, meningitis and respiratory disease with increasing frequency since 2004 [64, 65]. It is notable that 35% of 81 isolates in one of these studies belonged to phylogroup A, clonal complex 10 [64]. In European wild boars, which are assumed to be ancestors of domestic pigs in Europe [103], *E. coli* strains carry, on average, seven or more VAGs, with some strains carrying up to 16 VAGs [104]. Collectively, these observations suggest that *E. coli* phylogroup A and B1, at least those sourced from swine, carry multiple VAGs.

Contribution of food production animals to the evolution of pathogens and antimicrobial resistance

MDR *E. coli* carrying ARGs associated with mobile genetic elements and VAGs are released into the environment by food production animals via faecal effluent. In Australia the capacity for pig production to contribute to the evolution and dissemination of pathogens and ARGs is restricted compared to that of pig production systems in many other countries, due to a range of factors. Firstly, Australia has a large landmass that is surrounded by ocean preventing the

movement of animals from neighbouring countries. Secondly, importation of food animals into Australia has been restricted since the 1970s [105]. Thirdly, antibiotics such as fluoroquinolones cannot legally be administered to food animals and many others are restricted from use in food animal production [106, 107]. However, even in the restricted environment in Australia, phenotypic resistance to clinically-important antibiotics, including extended-spectrum cephalosporins and fluoroquinolones has been observed in *E. coli* which belong to globally disseminated *E. coli* lineages ST744, ST100 and ST1 [108]. Globally, genomic surveillance is needed to understand the relative contribution of food production animals to the complex web of interactions between microbiota and the mobile resistome; to provide baseline carriage rates for antimicrobial genes and VAGs; and to monitor the emergence of novel drug resistant pathogens [84, 109].

In summary, we report the first genomic study of commensal *E. coli* isolated from commercial pigs used for food consumption and provide data to inform assessment of potential risks pig commensal *E. coli* may pose to human health. Our results show that swine are a reservoir; i) for phylogroup A and B1 *E. coli* that carry VAGs, ii) the *sul3* gene; iii) class 1 integrons associated with IS*26*; and iv) *E. coli* lineages belonging to CC10. Our study has identified several new genetic signatures that may be used in tracking mobile antibiotic resistance genes.

4.2.7 References

1. **Poolman JT, Wacker M**. Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *J Infect Dis* 2016;213(1):6-13.
2. O'Neill J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. London: Wellcome Trust, UK Government2016.
3. **Herrero-Fresno A, Larsen I, Olsen JE**. Genetic relatedness of commensal Escherichia coli from nursery pigs in intensive pig production in Denmark and molecular characterization of genetically different strains. *J Appl Microbiol* 2015;119(2):342-353.
4. **Han S**. *Environmental Impacts of China's Pork Industry*. Wilson Center; 2014.
5. **Wang F-H, Ma W-Q, Dou Z-X, Ma L, Liu X-L et al.** The estimation of the production amount of animal manure and its environmental effect in China. *China Environmental Science* 2006;26(5):614-617.
6. **Johnson JR, Russo TA**. Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli. *Int J Med Microbiol* 2005;295(6-7):383-404.
7. **Johnson JR, Kuskowski MA, Smith K, O'Bryan TT, Tatini S**. Antimicrobial-resistant and extraintestinal pathogenic Escherichia coli in retail foods. *J Infect Dis* 2005;191(7):1040-1049.
8. **Varela AR, Manageiro V, Ferreira E, Guimaraes MA, da Costa PM et al.** Molecular evidence of the close relatedness of clinical, gull and wastewater isolates of quinolone-resistant Escherichia coli. *J Glob Antimicrob Resist* 2015;3(4):286-289.
9. **Manges AR**. Escherichia coli and urinary tract infections: the role of poultry-meat. *Clin Microbiol Infect* 2016;22(2):122-129.
10. **Nordstrom L, Liu CM, Price LB**. Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. *Front Microbiol* 2013;4:29.
11. **Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L et al.** Food reservoir for Escherichia coli causing urinary tract infections. *Emerg Infect Dis* 2010;16(1):88-95.
12. **Cordoni G, Woodward MJ, Wu H, Alanazi M, Wallis T et al.** Comparative genomics of European avian pathogenic E. Coli (APEC). *BMC Genomics* 2016;17(1):960.
13. **Levings RS, Djordjevic SP, Hall RM**. SGI2, a relative of Salmonella genomic island SGI1 with an independent origin. *Antimicrob Agents Chemother* 2008;52(7):2529-2537.
14. **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;84-85:11-19.
15. **Roy Chowdhury P, Charles IG, Djordjevic SP**. A role for Tn6029 in the evolution of the complex antibiotic resistance gene loci in genomic island 3 in enteroaggregative hemorrhagic Escherichia coli O104:H4. *PLoS One* 2015;10(2):e0115781.
16. **Venturini C, Beatson SA, Djordjevic SP, Walker MJ**. Multiple antibiotic resistance gene recruitment onto the enterohemorrhagic Escherichia coli virulence plasmid. *FASEB J* 2010;24(4):1160-1166.
17. **Venturini C, Hassan KA, Roy Chowdhury P, Paulsen IT, Walker MJ et al.** Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS26-related molecular signature isolated from different Escherichia coli pathotypes from different hosts. *PLoS One* 2013;8(11):e78862.

18. **Baker-Austin C, Wright MS, Stepanauskas R, McArthur JV**. Co-selection of antibiotic and metal resistance. *Trends in microbiology* 2006;14(4):176-182.
19. **Li LG, Xia Y, Zhang T**. Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. *Isme j* 2017;11(3):651-662.
20. **Ben W, Wang J, Pan X, Qiang Z**. Dissemination of antibiotic resistance genes and their potential removal by on-farm treatment processes in nine swine feedlots in Shandong Province, China. *Chemosphere* 2017;167:262-268.
21. **Zhou X, Qiao M, Wang FH, Zhu YG**. Use of commercial organic fertilizer increases the abundance of antibiotic resistance genes and antibiotics in soil. *Environmental science and pollution research international* 2017;24(1):701-710.
22. **Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM et al.** Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *Isme j* 2015;9(6):1269-1279.
23. **Liebert CA, Hall RM, Summers AO**. Transposon Tn21, flagship of the floating genome. *Microbiology and molecular biology reviews : MMBR* 1999;63(3):507-522.
24. **Cheng C, Sun J, Zheng F, Lu W, Yang Q et al.** New structures simultaneously harboring class 1 integron and ISCR1-linked resistance genes in multidrug-resistant Gram-negative bacteria. *BMC microbiology* 2016;16:71.
25. **Harmer CJ, Hall RM**. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* 2016;1(2).
26. **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(5):e01801-01814.
27. **Toleman MA, Bennett PM, Walsh TR**. ISCR elements: novel gene-capturing systems of the 21st century? *Microbiology and molecular biology reviews : MMBR* 2006;70(2):296-316.
28. **Canton R, Coque TM**. The CTX-M beta-lactamase pandemic. *Current opinion in microbiology* 2006;9(5):466-475.
29. **Dionisi AM, Lucarelli C, Owczarek S, Luzzi I, Villa L**. Characterization of the plasmid-borne quinolone resistance gene qnrB19 in Salmonella enterica serovar Typhimurium. *Antimicrob Agents Chemother* 2009;53(9):4019-4021.
30. **Toleman MA, Walsh TR**. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol Rev* 2011;35(5):912-935.
31. **Doublet B, Praud K, Weill FX, Cloeckaert A**. Association of IS26-composite transposons and complex In4-type integrons generates novel multidrug resistance loci in Salmonella genomic island 1. *J Antimicrob Chemother* 2009;63(2):282-289.
32. **Siebor E, Neuwirth C**. Emergence of Salmonella genomic island 1 (SGI1) among Proteus mirabilis clinical isolates in Dijon, France. *J Antimicrob Chemother* 2013;68(8):1750-1756.
33. **Mangat CS, Bekal S, Irwin RJ, Mulvey MR**. A Novel Hybrid Plasmid Carrying Multiple Antimicrobial Resistance and Virulence Genes in Salmonella enterica Serovar Dublin. *Antimicrob Agents Chemother* 2017;61(6).
34. **Porse A, Schonning K, Munck C, Sommer MO**. Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Molecular biology and evolution* 2016;33(11):2860-2873.

35. **Ahmed S, Olsen JE, Herrero-Fresno A**. The genetic diversity of commensal Escherichia coli strains isolated from non-antimicrobial treated pigs varies according to age group. *PLoS One* 2017;12(5):e0178623.
36. **Darling AE, Worden P, Chapman TA, Roy Chowdhury P, Charles IG et al.** The genome of Clostridium difficile 5.3. *Gut Pathog* 2014;6(1):4.
37. **Coil D, Jospin G, Darling AE**. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015;31(4):587-589.
38. **Bell S. M. PJN, Newton P., Nguyen T.T.** . Antibiotic Susceptibility Testing by the CDS Method: A Manual for Medical andVeterinary Laboratories. 2013;6.
39. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J et al.** BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
40. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al.** Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67(11):2640-2644.
41. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M**. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research* 2006;34(Database issue):D32-36.
42. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F**. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2015;53(8):2410-2426.
43. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS et al.** Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol* 2014;52(5):1501-1510.
44. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
45. **Langmead B, Salzberg SL**. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357-359.
46. **Milne I, Bayer M, Cardle L, Shaw P, Stephen G et al.** Tablet--next generation sequence assembly visualization. *Bioinformatics* 2010;26(3):401-402.
47. **Clermont O, Bonacorsi S, Bingen E**. Rapid and simple determination of the Escherichia coli phylogenetic group. *Applied and environmental microbiology* 2000;66(10):4555-4558.
48. **Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM et al.** PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014;2:e243.
49. **Price MN, Dehal PS, Arkin AP**. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5(3):e9490.
50. **Wyrsch E, Roy Chowdhury P, Abraham S, Santos J, Darling AE et al.** Comparative genomic analysis of a multiple antimicrobial resistant enterotoxigenic E. coli O157 lineage from Australian pigs. *BMC Genomics* 2015;16:165.
51. **Moran RA, Hall RM**. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. *Microbial drug resistance (Larchmont, NY)* 2017.
52. **Bednorz C, Oelgeschlager K, Kinnemann B, Hartmann S, Neumann K et al.** The broader context of antibiotic resistance: zinc feed supplementation of piglets increases the proportion of multi-resistant Escherichia coli in vivo. *Int J Med Microbiol* 2013;303(6-7):396-403.

53. **Cortes P, Blanc V, Mora A, Dahbi G, Blanco JE et al.** Isolation and characterization of potentially pathogenic antimicrobial-resistant Escherichia coli strains from chicken and pig farms in Spain. *Applied and environmental microbiology* 2010;76(9):2799-2805.
54. **Wang J, Gibbons JF, McGrath K, Bai L, Li F et al.** Molecular characterization of blaESBL-producing Escherichia coli cultured from pig farms in Ireland. *J Antimicrob Chemother* 2016.
55. **Alcala L, Alonso CA, Simon C, Gonzalez-Esteban C, Oros J et al.** Wild Birds, Frequent Carriers of Extended-Spectrum beta-Lactamase (ESBL) Producing Escherichia coli of CTX-M and SHV-12 Types. *Microbial ecology* 2016;72(4):861-869.
56. **Ho PL, Lo WU, Lai EL, Law PY, Leung SM et al.** Clonal diversity of CTX-M-producing, multidrug-resistant Escherichia coli from rodents. *Journal of medical microbiology* 2015;64(Pt 2):185-190.
57. **Liu X, Liu H, Li Y, Hao C**. High Prevalence of beta-lactamase and Plasmid-Mediated Quinolone Resistance Genes in Extended-Spectrum Cephalosporin-Resistant Escherichia coli from Dogs in Shaanxi, China. *Front Microbiol* 2016;7:1843.
58. **Manges AR, Harel J, Masson L, Edens TJ, Portt A et al.** Multilocus sequence typing and virulence gene profiles associated with Escherichia coli from human and animal sources. *Foodborne Pathog Dis* 2015;12(4):302-310.
59. **Rodrigues C, Machado E, Peixe L, Novais A**. IncI1/ST3 and IncN/ST1 plasmids drive the spread of blaTEM-52 and blaCTX-M-1/-32 in diverse Escherichia coli clones from different piggeries. *J Antimicrob Chemother* 2013;68(10):2245-2248.
60. **Shabana, II, Zaraket H, Suzuki H**. Molecular studies on diarrhea-associated Escherichia coli isolated from humans and animals in Egypt. *Vet Microbiol* 2013;167(3-4):532-539.
61. **Trobos M, Christensen H, Sunde M, Nordentoft S, Agerso Y et al.** Characterization of sulphonamide-resistant Escherichia coli using comparison of sul2 gene sequences and multilocus sequence typing. *Microbiology* 2009;155(Pt 3):831-836.
62. **Guiral E, Mendez-Arancibia E, Soto SM, Salvador P, Fabrega A et al.** CTX-M-15-producing enteroaggregative Escherichia coli as cause of travelers' diarrhea. *Emerg Infect Dis* 2011;17(10):1950-1953.
63. **Reuland EA, Overdevest IT, Al Naiemi N, Kalpoe JS, Rijnsburger MC et al.** High prevalence of ESBL-producing Enterobacteriaceae carriage in Dutch community patients with gastrointestinal complaints. *Clin Microbiol Infect* 2013;19(6):542-549.
64. **Ding Y, Tang X, Lu P, Wu B, Xu Z et al.** Clonal analysis and virulent traits of pathogenic extraintestinal Escherichia coli isolates from swine in China. *BMC veterinary research* 2012;8:140.
65. **Tan C, Tang X, Zhang X, Ding Y, Zhao Z et al.** Serotypes and virulence genes of extraintestinal pathogenic Escherichia coli isolates from diseased pigs in China. *Veterinary journal (London, England : 1997)* 2012;192(3):483-488.
66. **Giufre M, Graziani C, Accogli M, Luzzi I, Busani L et al.** Escherichia coli of human and avian origin: detection of clonal groups associated with fluoroquinolone and multidrug resistance in Italy. *J Antimicrob Chemother* 2012;67(4):860-867.
67. **Salvador E, Wagenlehner F, Kohler CD, Mellmann A, Hacker J et al.** Comparison of asymptomatic bacteriuria Escherichia coli isolates from healthy individuals versus those from hospital patients shows that long-term bladder

colonization selects for attenuated virulence phenotypes. *Infect Immun* 2012;80(2):668-678.
68. **Usein CR, Papagheorghe R, Oprea M, Condei M, Straut M**. Molecular characterization of bacteremic Escherichia coli isolates in Romania. *Folia microbiologica* 2016;61(3):221-226.
69. **Pires J, Kuenzli E, Kasraian S, Tinguely R, Furrer H et al.** Polyclonal Intestinal Colonization with Extended-Spectrum Cephalosporin-Resistant Enterobacteriaceae upon Traveling to India. *Front Microbiol* 2016;7:1069.
70. **Tagg KA, Ginn AN, Partridge SR, Iredell JR**. MALDI-TOF Mass Spectrometry for Multilocus Sequence Typing of Escherichia coli Reveals Diversity among Isolates Carrying blaCMY(-)(2)-Like Genes. *PLoS One* 2015;10(11):e0143446.
71. **Alghoribi MF, Gibreel TM, Farnham G, Al Johani SM, Balkhy HH et al.** Antibiotic-resistant ST38, ST131 and ST405 strains are the leading uropathogenic Escherichia coli clones in Riyadh, Saudi Arabia. *J Antimicrob Chemother* 2015;70(10):2757-2762.
72. **Toval F, Kohler CD, Vogel U, Wagenlehner F, Mellmann A et al.** Characterization of Escherichia coli isolates from hospital inpatients or outpatients with urinary tract infection. *J Clin Microbiol* 2014;52(2):407-418.
73. **Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J, Voets GM, van den Munckhof MP et al.** Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clin Microbiol Infect* 2011;17(6):873-880.
74. **Moran RA, Anantham S, Pinyon JL, Hall RM**. Plasmids in antibiotic susceptible and antibiotic resistant commensal Escherichia coli from healthy Australian adults. *Plasmid* 2015;80:24-31.
75. **Antunes P, Machado J, Peixe L**. Dissemination of sul3-containing elements linked to class 1 integrons with an unusual 3' conserved sequence region among Salmonella isolates. *Antimicrob Agents Chemother* 2007;51(4):1545-1548.
76. **Curiao T, Canton R, Garcillan-Barcia MP, de la Cruz F, Baquero F et al.** Association of composite IS26-sul3 elements with highly transmissible IncI1 plasmids in extended-spectrum-beta-lactamase-producing Escherichia coli clones from humans. *Antimicrob Agents Chemother* 2011;55(5):2451-2457.
77. **Liu J, Keelan P, Bennett PM, Enne VI**. Characterization of a novel macrolide efflux gene, mef(B), found linked to sul3 in porcine Escherichia coli. *J Antimicrob Chemother* 2009;63(3):423-426.
78. **Gundogdu A, Long YB, Vollmerhausen TL, Katouli M**. Antimicrobial resistance and distribution of sul genes and integron-associated intI genes among uropathogenic Escherichia coli in Queensland, Australia. *Journal of medical microbiology* 2011;60(Pt 11):1633-1642.
79. **Bischoff KM, White DG, Hume ME, Poole TL, Nisbet DJ**. The chloramphenicol resistance gene cmlA is disseminated on transferable plasmids that confer multiple-drug resistance in swine Escherichia coli. *FEMS Microbiol Lett* 2005;243(1):285-291.
80. **Partridge SR, Tsafnat G, Coiera E, Iredell JR**. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol Rev* 2009;33(4):757-784.
81. **Saenz Y, Vinue L, Ruiz E, Somalo S, Martinez S et al.** Class 1 integrons lacking qacEDelta1 and sul1 genes in Escherichia coli isolates of food, animal and human origins. *Vet Microbiol* 2010;144(3-4):493-497.

82. **Sunde M, Solheim H, Slettemeas JS**. Genetic linkage between class 1 integrons with the dfrA12-orfF-aadA2 cassette array and sul3 in Escherichia coli. *Vet Microbiol* 2008;130(3-4):422-425.
83. **Alonso CA, Michael GB, Li J, Somalo S, Simon C et al.** Analysis of blaSHV-12-carrying Escherichia coli clones and plasmids from human, animal and food sources. *J Antimicrob Chemother* 2017;72(6):1589-1596.
84. **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.
85. **Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP et al.** Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in Escherichia coli of human and animal origin. *PLoS One* 2010;5(9):e12754.
86. **Levings RS, Lightfoot D, Partridge SR, Hall RM, Djordjevic SP**. 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. *Journal of bacteriology* 2005;187(13):4401-4409.
87. **Cain AK, Liu X, Djordjevic SP, Hall RM**. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant Salmonella enterica serovar Typhimurium from Australian animals. *Microbial drug resistance (Larchmont, NY)* 2010;16(3):197-202.
88. **Reid CJ, Roy Chowdhury P, Djordjevic SP**. Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. *Plasmid* 2015;80:127-137.
89. **Bleibtreu A, Gros PA, Laouenan C, Clermont O, Le Nagard H et al.** Fitness, stress resistance, and extraintestinal virulence in Escherichia coli. *Infect Immun* 2013;81(8):2733-2742.
90. **Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL et al.** Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. *J Infect Dis* 2013;207(6):919-928.
91. **Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE et al.** Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. *MBio* 2016;7(2):e02162.
92. **Zhu Ge X, Jiang J, Pan Z, Hu L, Wang S et al.** Comparative genomic analysis shows that avian pathogenic Escherichia coli isolate IMT5155 (O2:K1:H5; ST complex 95, ST140) shares close relationship with ST95 APEC O1:K1 and human ExPEC O18:K1 strains. *PLoS One* 2014;9(11):e112048.
93. **Mora A, Lopez C, Herrera A, Viso S, Mamani R et al.** Emerging avian pathogenic Escherichia coli strains belonging to clonal groups O111:H4-D-ST2085 and O111:H4-D-ST117 with high virulence-gene content and zoonotic potential. *Vet Microbiol* 2012;156(3-4):347-352.
94. **Schierack P, Walk N, Ewers C, Wilking H, Steinruck H et al.** ExPEC-typical virulence-associated genes correlate with successful colonization by intestinal E. coli in a small piglet group. *Environ Microbiol* 2008;10(7):1742-1751.
95. **Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V et al.** Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proc Natl Acad Sci U S A* 2006;103(15):5977-5982.

96. **Srinivasan U, Foxman B, Marrs CF**. Identification of a gene encoding heat-resistant agglutinin in Escherichia coli as a putative virulence factor in urinary tract infection. *J Clin Microbiol* 2003;41(1):285-289.
97. **Johnson JR, Jelacic S, Schoening LM, Clabots C, Shaikh N et al.** The IrgA homologue adhesin Iha is an Escherichia coli virulence factor in murine urinary tract infection. *Infect Immun* 2005;73(2):965-971.
98. **Kaper JB, Nataro JP, Mobley HL**. Pathogenic Escherichia coli. *Nat Rev Microbiol* 2004;2(2):123-140.
99. **Ewers C, Li G, Wilking H, Kiessling S, Alt K et al.** Avian pathogenic, uropathogenic, and newborn meningitis-causing Escherichia coli: how closely related are they? *Int J Med Microbiol* 2007;297(3):163-176.
100. **Johnson JR, Stell AL**. Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* 2000;181(1):261-272.
101. **Dixit SM, Gordon DM, Wu XY, Chapman T, Kailasapathy K et al.** Diversity analysis of commensal porcine Escherichia coli - associations between genotypes and habitat in the porcine gastrointestinal tract. *Microbiology* 2004;150(Pt 6):1735-1740.
102. **Schierack P, Walk N, Reiter K, Weyrauch KD, Wieler LH**. Composition of intestinal Enterobacteriaceae populations of healthy domestic pigs. *Microbiology* 2007;153(Pt 11):3830-3837.
103. **Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E et al.** Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science (New York, NY)* 2005;307(5715):1618-1621.
104. **Romer A, Wieler LH, Schierack P**. Analyses of intestinal commensal Escherichia coli strains from wild boars suggest adaptation to conventional pig production conditions. *Vet Microbiol* 2012;161(1-2):122-129.
105. **Turner A**. Quarantine, exports and animal disease in Australia 1901-2010. *Australian veterinary journal* 2011;89(9):366-371.
106. **Cheng AC, Turnidge J, Collignon P, Looke D, Barton M et al.** Control of fluoroquinolone resistance through successful regulation, Australia. *Emerg Infect Dis* 2012;18(9):1453-1460.
107. **Jordan D, Chin JJ, Fahy VA, Barton MD, Smith MG et al.** Antimicrobial use in the Australian pig industry: results of a national survey. *Australian veterinary journal* 2009;87(6):222-229.
108. **Abraham S, Jordan D, Wong HS, Johnson JR, Toleman MA et al.** First detection of extended-spectrum cephalosporin- and fluoroquinolone-resistant Escherichia coli in Australian food-producing animals. *J Glob Antimicrob Resist* 2015;3(4):273-277.
109. **Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM et al.** Genomic Microbial Epidemiology Is Needed to Comprehend the Global Problem of Antibiotic Resistance and to Improve Pathogen Diagnosis. *Front Microbiol* 2016;7:843.

4.2.8 Data Bibliography

1. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S*, et al.*** Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*. 2012;67(11):2640-4.

2. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M**. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research*. 2006;34(Database issue):D32-6.
3. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F**. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol*. 2015;53(8):2410-26.
4. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS*, et al.*** Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol*. 2014;52(5):1501-10.
5. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O*, et al.*** In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014;58(7):3895-903.

4.2.9 Figures

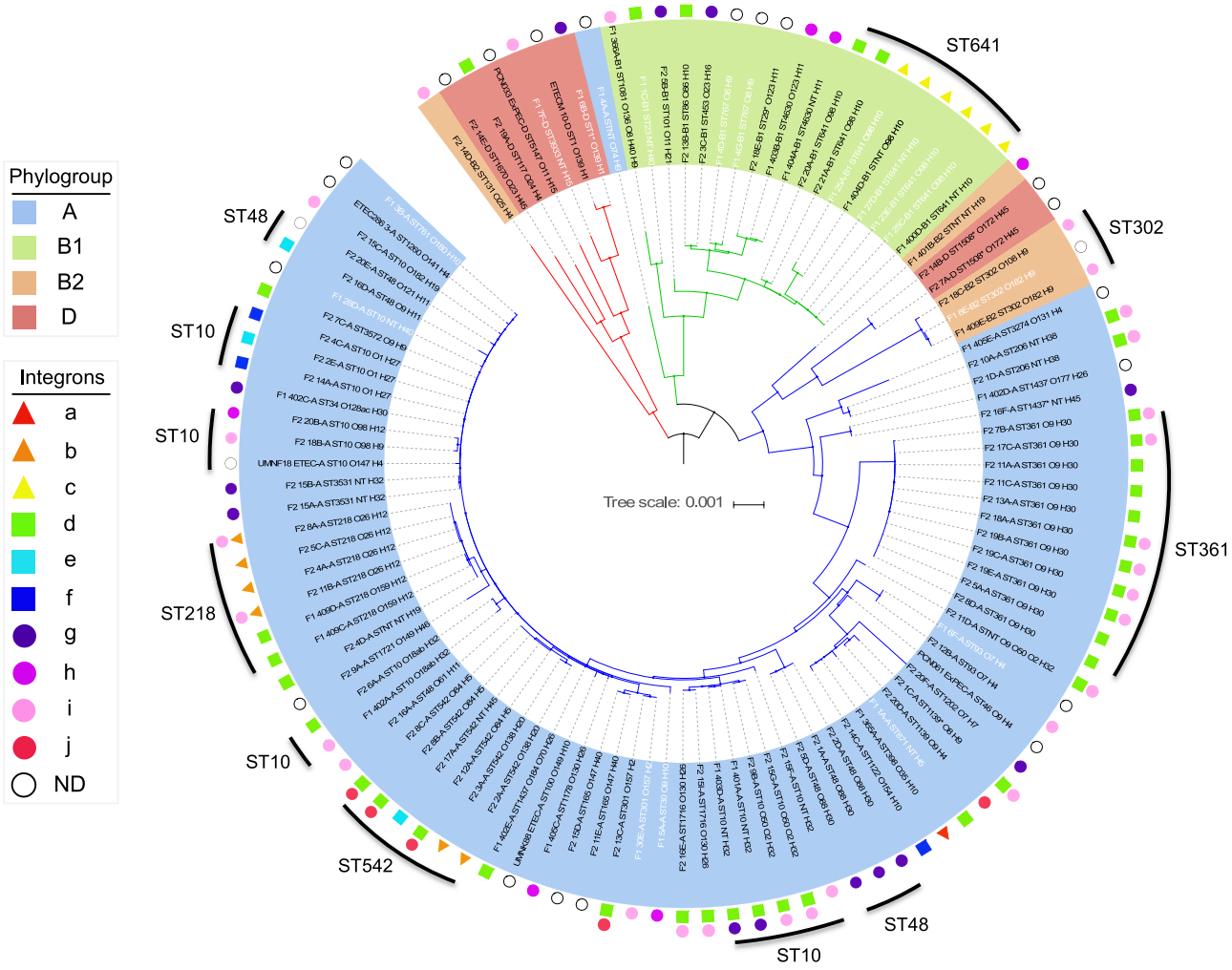


Figure 4.1

A mid-point rooted, maximum-likelihood phylogenetic tree inferred using PhyloSift v1.0.1, FastTree2, FigTree v1.4.2 and iTOL. The tree contains all 103 pig *E. coli* isolates sequenced in this study, two porcine ETEC strains and four reference pig-sourced sequences. Strains isolated from pigs with diarrhoea are in white and ETEC and reference strains are in bold. Branches are coloured by clade (Clade 1: red, Clade 2: green, Clade 3: Blue). Shading over tip labels indicates phylogroup (A-blue, B1-green, B2-orange or D-red) Tip labels also contain MLST and serotype. Asterisks indicate single-locus variants of a given sequence type. The tree scale shows the distance for 1 amino acid substitution per 1000 sites in the analysis. Clusters of the 7 most common sequence types have been marked as an outer line. Integrons shown in Fig. 3 are annotated by shapes indicating presence of *sul1* (triangles), IS*26*-truncated 3`-*CS* (squares) and *sul3* (circles). Integrons (a-j) are coloured by red, orange, yellow, green, aqua, blue, purple, magenta, pink and crimson respectively. Strains that were *intI1* positive but were not characterised are annotated with a blank circle. Integrons were not determined for reference genomes used in the analysis.

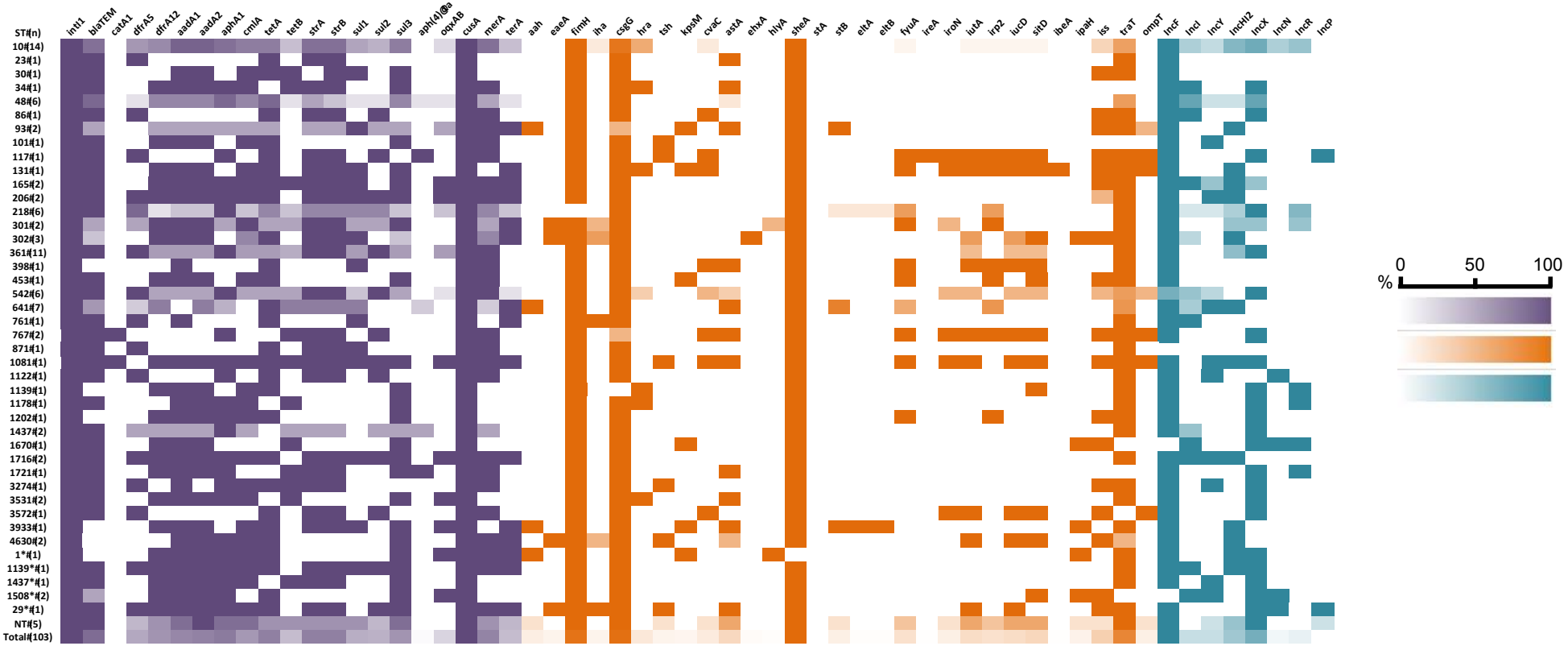


Figure 4.2

Heat map depicting carriage of antimicrobial resistance genes (aqua), virulence-associated genes (orange) and plasmid incompatibility groups (purple) by sequence type. Darker colour indicates high carriage amongst a given sequence type, lighter colour indicates lower carriage and white indicates no carriage. For full screening data see Table S4.3-S4.5.

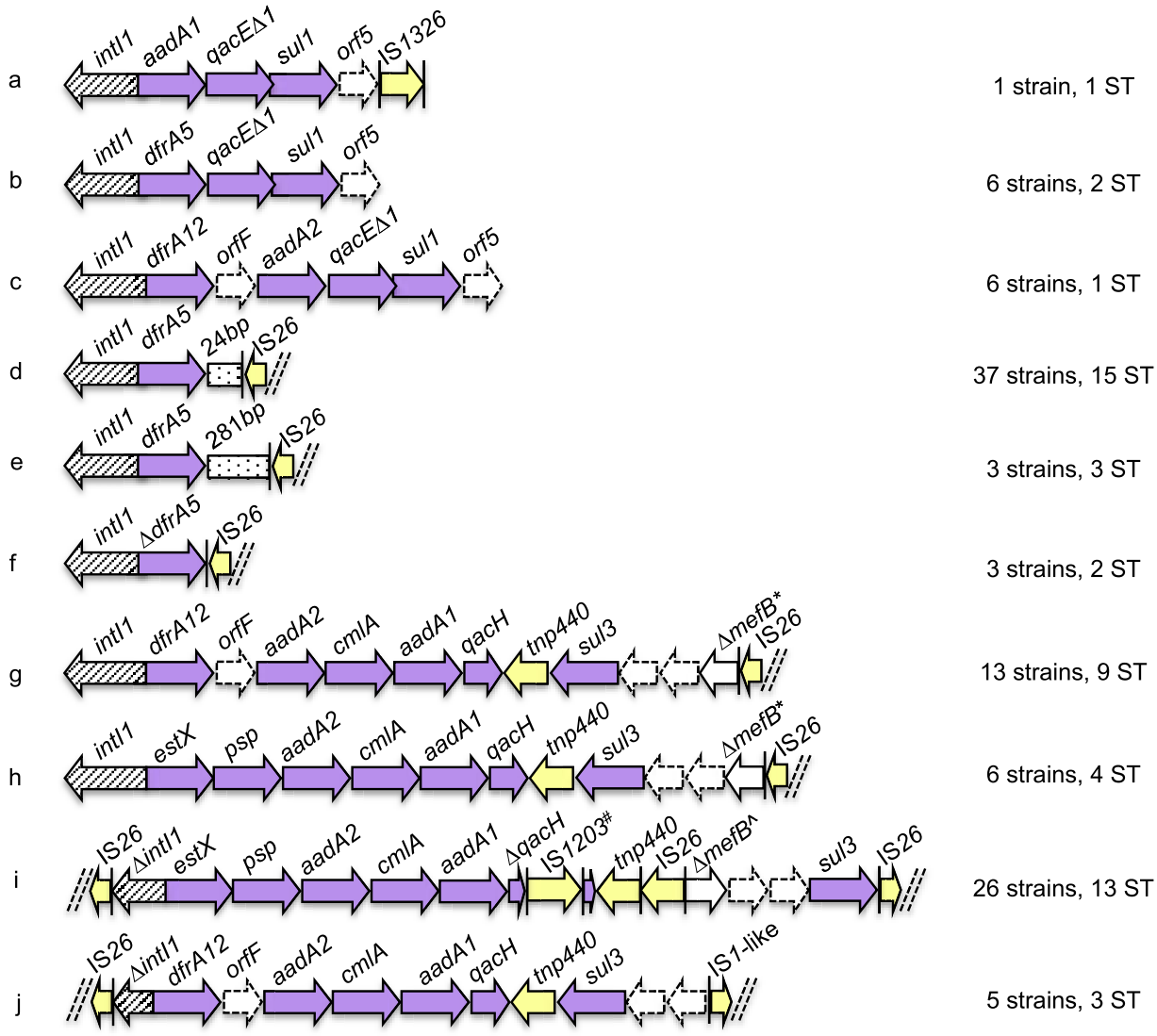


Figure 4.3

Schematic diagram (not to scale) of integrons within porcine strains that were sequenced. Arrows represent ORFs. Arrows with broken lines indicate hypothetical proteins. Vertical bars represent inverted repeats. Dashed double diagonal lines represent sequence scaffold breaks. Intergenic sequences are not shown. Antimicrobial resistance genes (purple) and IS/transposable elements (yellow), are colour coded.

*indicates 260 bp of *mefB* remaining.

^ indicates 111 bp of *mefB* remaining.

IS*1203*-like

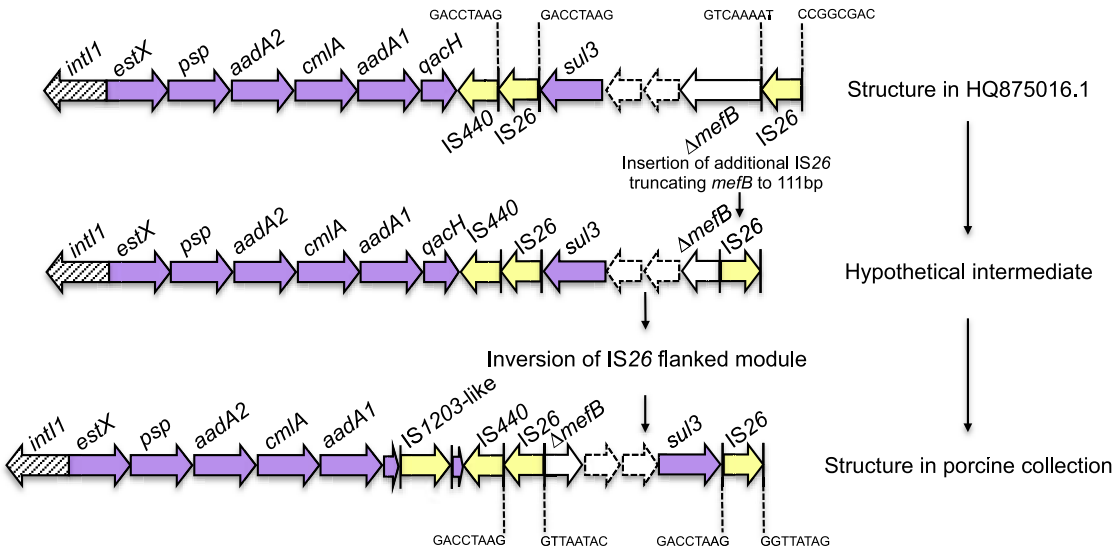


Figure 4.4

Schematic diagram (not to scale) of proposed evolutionary pathway to the *sul3-ΔmefB* arrangement shown in Figure 3 (i). IS*26* 8 bp direct repeats are annotated.

Chapter 5: Australian porcine clonal complex 10 (CC10) *Escherichia coli* belong to multiple sublineages of a highly diverse global CC10 phylogeny

5.1 Declaration

Cameron J. Reid, Matthew Z. DeMaere, Steven P. Djordjevic*

The i3 institute, University of Technology Sydney, Ultimo, NSW 2007, Australia
*corresponding author

Published in Microbial Genomics, October 2018, doi: 10.1099/mgen.0.000225

Author contribution statement

This research publication was conceived by C.J.R. and S.P.D. All data was generated and analysed by C.J.R. M.Z.D. assisted with generation and interpretation of MDS data. The manuscript was written by C.J.R. and edited by S.P.D.

5.2 Australian porcine CC10 *E. coli* belong to multiple sublineages of a highly diverse global CC10 phylogeny

Following the identification of CC10 as the dominant group among MDR porcine commensal *E. coli* and having established their carriage of multiple integrons and ARGs, it was decided to compare them to a global collection of CC10 strains from various sources. It was hoped this would shed some light on the diversity and origins of CC10 that colonise pigs as well as provide a global snapshot of CC10 for the purposes of future genomic epidemiology concerning this widespread clonal group. This chapter primarily addresses Aim 2 of the thesis. Please view figures as PDFs attached to this thesis.

5.2.1 Abstract

We recently identified CC10 *Escherichia coli* as the predominant clonal group in two populations of healthy Australian food production pigs. CC10 are highly successful, colonising humans, food production animals, fresh produce and environmental niches. Furthermore, *E. coli* within CC10 are frequently drug resistant and increasingly reported as human and animal extra-intestinal pathogens. In order to develop a high-resolution global phylogeny and determine

the repertoire of antimicrobial resistance genes, virulence-associated genes and plasmid types within this clonal group, we downloaded 228 publicly available CC10 short read genome sequences for comparison with 20 porcine CC10 we have previously described. Core genome SNP phylogeny revealed a highly diverse global phylogeny consisting of multiple lineages that did not cluster by geography or source of isolation. Australian porcine strains belonged to several of these divergent lineages indicating CC10 is present in these animals due to multiple colonisation events. Differences in resistance gene and plasmid carriage between porcine strains and the global collection highlighted the role of lateral gene transfer in the evolution of CC10 strains. Virulence profiles typical of extra-intestinal pathogenic *E. coli* were present in both Australian porcine strains and the broader collection. As both the core phylogeny and accessory gene characteristics appeared unrelated to geography or source of isolation it is likely that the global expansion of CC10 is not a recent event and may be associated with faecal carriage in humans.

5.2.2 Data Summary

1. All R code used in this study, SNP tree file and a fasta database of additional screening genes have been deposited at https://github.com/CJREID/CC10_supporting_data

2. Whole genome sequenced short reads for porcine strains in this study are available at Enterobase and ENA; accession numbers are listed in Table S5.1.

5.2.3 Introduction

Escherichia coli is both a successful commensal and a serious pathogen affecting human and animal health and is the most frequently isolated Gram negative pathogen impacting human health [1]. Multidrug resistant (MDR; resistant to three or more classes of antimicrobials) infections are forecast to cause 10 million deaths per year by 2050 and it is expected that *E. coli* will be responsible for 30% of fatalities and 40% of projected economic losses that arise as a consequence [2]. Pathogenic *E. coli* may infect intestinal (InPEC; intestinal pathogenic *E. coli*) or extra-intestinal sites (ExPEC; extra-intestinal pathogenic *E. coli*)[3] and antimicrobial resistance (AMR) often complicates treatment, increasing rates of

morbidity and mortality [4]. Severe outbreaks of drug-resistant InPEC disease such as O104:H4 in 2011 are well documented as well as global dissemination of resistant ExPEC clones such as ST131 [5-7]. Currently, responses to such events are reactive and come after significant financial and human cost is incurred. It would therefore be beneficial to be able to identify, monitor and track populations of *E. coli* that pose a threat to human health so that risks can be predicted and strategies implemented to mitigate impact. To this end, the emerging field of genomic epidemiology is critical as it allows the highest resolution of microbial population structure and genetic determinants of virulence and AMR. By developing global databases of genomic, phenotypic, spatial and temporal information for *E. coli,* predictive disease management is a distinct possibility for the near future [8, 9].

There is currently much debate about the contribution of different reservoirs of *E. coli* to disease and antimicrobial resistance carriage in humans [9-13]. Food production animals and associated retail meats have been widely investigated and genetic similarities documented, however the data is often limited to MLST types and PCR identification of a select number of genes [14, 15]. Furthermore, the sample sizes used are a limiting factor in the significance of the conclusions that may be reached. Nevertheless, these studies provide a good starting point for further genomic investigation.

We recently described the phylogeny, virulence-associated gene (VAG) and antimicrobial resistance gene (ARG) carriage in a collection of 103 *E. coli* genome sequences derived from the faeces of healthy Australian pigs [16]. This study identified clonal complex 10 (CC10) as the predominant lineage within the collection. These CC10 strains were phylogenetically diverse and carried multiple ARGs and VAGs associated with ExPEC. CC10 is globally reported as a resident of the intestinal tract of humans, food production animals, companion animals and wild animals [14, 17]. It has also been identified in retail meats and plant-based foods as well as wastewater, rivers and urban streams [18, 19]. Furthermore, CC10 can cause extra-intestinal disease in pigs, dogs and humans [14, 17, 20]. Antimicrobial resistance, including ESBL carriage is also widely reported [19, 21].

These observations suggest a clonal complex with broad fitness characteristics, a wide host-range, pathogenic potential and a variety of antimicrobial resistance traits. Despite the wealth of literature reporting CC10, the global population structure and diversity of VAGs and ARGs remains unknown.

The aim of this study was firstly to examine a global collection of CC10 *E. coli* genome sequences to determine the population structure and the diversity of VAGs, ARGs and plasmid replicons. Secondly we aimed to determine how Australian porcine CC10 *E. coli* relate to the global phylogeny and to compare VAG, ARG and plasmid replicon carriage to the existing global collection.

5.2.4 Methods

Genome sequences used in this study

The 20 Australian porcine faecal-derived *E. coli* CC10 strains included in this analysis were whole genome sequenced using a modified Nextera protocol and Illumina HiSeq platform as previously described [16]. These strains retain their original names from the previous publication and are preceded by 'F2_'. The 20 strains described in this paper are available as short reads at http://enterobase.warwick.ac.uk/ and the European Nucleotide Archive. All accession numbers are listed in Table S5.1.

Publicly available *E. coli* CC10 Illumina short reads were downloaded from Enterobase (http://enterobase.warwick.ac.uk/; accessed 22/11/17). The database was queried for CC10 and a summary spreadsheet was downloaded in order to select sequences from well-defined sources, with desired metadata for which full short reads were available. Accepted sources were: i) animal including faeces (pig, cattle, poultry, horse, dog); ii) food (pig, cattle, poultry, dairy cheese, plant); iii) human including faeces, urine and blood and; iv) environment including soil and wastewater. Furthermore, metadata was required for year of isolation, country and continent of origin. Sequences derived from laboratory strains were excluded, as were sequences with ambiguous or contradictory source details. This spreadsheet was then used to query the NCBI SRA archive and download read sets using the download_enterobase_SRA_reads.sh script available at https://github.com/bogemad/snp_phylogeny. Enterobase sequences were named

for analysis by removing the prefix 'ESC_' and suffix 'AA' from their 'Uberstrain' Accession numbers (Table S5.1). Preliminary analysis indicated a number of clones were present in the collection; therefore we selected a single representative strain in cases where groups of sequences were separated by ≤ 3 SNPs and had identical plasmid replicon and resistance gene profiles. The final collection numbered 248 sequences, comprising 228 from Enterobase and 20 from our previous study.

Phylogenetic Analysis

Snippy v4.0.2 (https://github.com/tseemann/snippy) was used with default parameters to map short reads from the 248 CC10 strains, as well as phylogroup A non-ST10 strain HS (gb|CP000802.1) to ST10 reference sequence K12-MG1655 (gb|U00096.3). A core genome alignment was then generated using the snippy-core function. This function generates two alignments, a 'full core' alignment of all reads to the reference genome and a 'core SNP' alignment consisting only of SNP sites present in all genomes, ignoring insertion and deletion variant types. Both of these alignments were used to generate maximum-likelihood trees. Files available as supporting data are hereafter named in parentheses. The full alignment was cleaned using the snippy-clean_full_aln function and filtered for recombination using default settings with Gubbins v2.3.4 [22] resulting in an alignment of 79,039 sites (full.core.clean.gubbins.aln). Variable positions present in all strains were then identified with SNP-sites v2.4.0 [23]. The final alignment consisted of 4515 SNP positions (full.core.clean.gubbins.snpsites.aln). The core SNP alignment generated by snippy (snippycore.aln) consisted of 72,136 sites. FastTree2 v2.1.10 [24] was used to generate maximum-likelihood phylogenetic trees under a GTR nucleotide substitution model with default settings for both of these alignments (full.core.clean.gubbins.snpsites.tree and snippycore.tree). The trees were visualised with metadata in iTOL 4.2.3 [25]. The 'full core' tree was also rooted in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and then visualised alongside gene-screening heatmaps in R (version 3.3.1) with ggtree 3.6 (full.core.clean.gubbins.snpsites.rooted.tree) [26]. Pairwise SNPs between all strains were extracted from the recombination filtered 'full core' alignment using the snp_phylo_utils script available at https://github.com/bogemad/snp_phylogeny. All trees, alignments used to

generate them, summary statistics from snippy-core and pairwise SNP table are available at https://github.com/CJREID/CC10_supporting_data.

Gene screening

Antimicrobial resistance genes, virulence-associated genes, plasmid replicon genes and OH antigen genes were identified using ARIBA (version 2.10.1) [27] with ResFinder, PlasmidFinder, VirulenceFinder and SerotypeFinder databases available from http://www.genomicepidemiology.org/. [Data references 1-4]. Pasteur MLST types were also determined with ARIBA's built in MLST typing function. An additional custom database of further virulence and resistance-associated genes not present in the aforementioned databases was also used. This database is available at https://github.com/CJREID/CC10_supporting_data. Gene presence or absence was then visualised with the phylogenetic tree in ggtree [26].

Multidimensional scaling analysis

In order to determine if virulence or resistance and plasmid gene carriage was related to continent of isolation, sequence type or origin of CC10 strains, we conducted a non-metric multidimensional scaling analysis in R using MASS and vegan packages. Briefly, the Jaccard index was used to produce a pairwise distance matrix for absence/presence profiles of virulence and combined resistance and plasmid replicon genes for each strain. Multi-dimensional scaling was performed using the methods MASS::isoMDS [28] in combination with vegan::initMDS and vegan::postMDS [29] with default settings. First, metric scaling was assumed to generate a baseline solution in two dimensions before iterating the same process with initMDS to reduce the stress value. postMDS was then used to standardise the final configurations for ease of interpretation. ggplot2 was used to plot the ordinations with clusters keyed separately on origin, sequence type and continent of isolation. A normal distribution was assumed to infer 95% confidence interval (CI) ellipses. All R code used in this study is available at https://github.com/CJREID/CC10_supporting_data.

R Package Versions

APE 5.0, MASS 7.3, vegan 2.4-5, dplyr 0.7.1, reshape2 1.4.2, grid 3.3.1, ggtree 1.6.11, ggplot2 2.2.1

5.2.5 Results and Discussion

The collection of CC10 sequences (n=248) encompassed eight sequence types (Achtman) from five continents. Pasteur sequence types were also determined, however concordance between the two schemes was highly variable (Table S5.1). Sequences were classified as being of food, animal, environment or human origin (Table 5.1). Strains were further classified into 16 sources, though four of these, for which there were less than 10 representatives, were grouped as 'Other' for analysis (Fig. 5.1, Table S5.1). The collection was heavily weighted towards North American samples however the distribution between animal, food and human sources was fairly even. Environmental samples were scarce. Isolation dates mostly ranged from 1979-2017, though one sequence was derived from a strain isolated in 1895. 122 serotypes were predicted and ten of these were O-non-typeable (Table S5.1).

Global phylogeny of E. coli CC10

In order to understand the population structure of our collection of CC10, a whole genome alignment core SNP phylogeny was generated using the completed genome of ST10 *E. coli* K12-MG1655 as a reference and the complete genome of ST46 phylogroup A *E. coli* HS as an out-group strain. The average number of bases aligned to the reference in the full core alignment was 4,272,312/4,643,559 (92%). Recombination filtering reduced this alignment to 79,039 bases and SNP identification resulted in a final alignment of 4,515 variable sites present in all strains. A maximum-likelihood tree was built from this alignment and the tree comprised four well-supported major clades. Clade 1 comprised only ST48 strains. This clade was separated from clade 2 by approximately 450 core SNPs. Interestingly, another cluster of ST48 strains was present in clade 4, differing from ST10 strains by approximately 90 core SNPs. This suggests that ST48 comprises two separate evolutionary lineages that may have been separated by geographical or host-isolation for some time before disseminating again. A maximum-likelihood

tree generated with the snippy-core core SNP alignment also supported this topology, albeit with some rearrangements in clade 4 (Fig. S5.1). This high-resolution approach supports previous work on these strains, in which Phylosift analysis also separated ST48 strains in the context of a wider collection of sequence types [27]. We are unaware of any other phylogenomic studies that demonstrate such a split between members of the same sequence type. However, intra-ST diversity, divergence and instances where core genome alignment does not strictly follow preceding typing schemes are reported [7, 30, 31]. It is difficult to compare results between studies due to different methodologies, the effect of different reference genomes and the number of strains included in the analysis. Nonetheless, this result supports whole genome core SNP phylogeny together with MLST to accurately resolve clonal groups [30].

ST218 strains formed the second clade whilst a single ST215 formed a third clade. Clade 4 was the largest and most diverse with respect to sequence types, containing ST10, ST34, ST43, ST44, ST48 and ST167. Major splits in the phylogeny were well supported, however some poor node confidence values were present within the sub-clades of clade 4, usually at splits between ST10 strains. The major lineages did not cluster sequences based on continent of isolation, origin or source (Fig. 5.1). There were numerous examples of sequences from disparate sources, origins and continents being closely related. This data supports a diverse clonal group consisting of multiple lineages with broad fitness characteristics that is globally dispersed and capable of inhabiting a wide variety of niches.

Australian porcine E. coli CC10 sequences in the context of global phylogeny

Thirteen Australian porcine sequences belonged to clade 4 (n=13) while the remaining seven ST48 and ST218 sequences belonged to clades 1 and 2 respectively. This indicates that CC10 lineages have been introduced to Australian pigs multiple times. Within clade 4, a number of Australian porcine sequences clustered with temporally and geographically unrelated sequences. Sequences that clustered with Australian porcine strains included human blood, urine and faecal, turkey meat, chicken faecal and chicken meat. This observation supports a growing body of literature that suggest CC10 *E. coli* are capable of colonizing a broad range

of hosts in both commensal and pathogenic capacities and does not preclude the possibility that porcine-origin CC10 are capable of causing extra-intestinal infection [14, 32].

Interestingly, Australian porcine sequences were mostly present on different sub-clades to the 14 other porcine faecal strains in the collection indicating divergence and diversity among CC10 strains that colonise pigs. However, two examples of closely related pig-derived strains were present in clade 4. Porcine faecal strain EA2788, isolated in 1979 in the USA, was separated by 25 core SNPs from its closest Australian relative F2_2E. Strain AA8187, isolated in the USA in 2009 from pig meat was yet more closely related to F2_2E, separated by only 9 core SNPs.

These three sequences possessed similar virulence-associated gene profiles, however their resistance genes profiles were vastly different. The Australian strain carried 15 ARGs compared to 7 in AA8187 and 3 in EA2788. Furthermore F2_2E carried a full suite of IncF replicons (FII, FIA, FIB, FIC), whilst AA8187 carried an IncFIA and an IncHI1 replicon and EA2788 carried an IncY replicon. These observations indicate that some CC10 strains are capable of persisting for long periods of time in association with the porcine gut. Whilst the actual genetic determinants of fitness in the porcine gut are not known, the phylogenetic similarity between these three strains likely reflects carriage of similar if not identical fitness factors. This persistence has previously been described between sows and their piglets and is likely to persist on a global scale due to the common ancestry of all domestic pig breeds [33, 34]. It would be interesting to see if the similarity between these Australian and American porcine strains is reflected in CC10 isolates from China, which are underrepresented in this collection. China is the world's largest producer of pork [35] and transfer of similar strains is likely to have occurred through trade of meat products and human travel between the two countries. A broader, globally sourced collection of porcine CC10 sequences would be useful to characterise dominant strains and determine the underlying genetic basis of their fitness. In contrast to the conservation of closely related strains, it is evident among pig-associated strains that variability still arises in the form of

accessory gene content such as plasmids and ARGs, likely due to to differential antimicrobial selective pressures between pork production systems worldwide.

CC10 carry a wide variety of virulence-associated genes

We identified a total of 110 virulence-associated genes within the collection. Strains carried between 2 and 34 virulence-associated genes, with an average of 10 VAGs per strain, revealing a wide variety of VAG profiles. The average number of VAGs per strain by source was highest in humans (n=12), followed by food and environmental (n=10) and animals (n=7). No clear relationship between the VAG profiles and source or geography was evident when gene presence/absence was mapped alongside the SNP phylogeny (Fig. S5.2). Some closely related strains exhibited similar virulence profiles however they were rarely identical. We therefore conducted a non-metric MDS analysis of 194 unique virulence profiles by continent of isolation, sequence type and origin. This analysis did not separate strains based on any of these characteristics, suggesting that virulence profiles are detached from these factors (Fig. 5.2(a-c), Fig. S5.2) It should be noted that small sample sizes for some groups may obscure the true distributions inferred by the ellipses, however this was unable to be avoided due to the limitations of publicly available sequence data. Overall, these observations highlight significant diversity in the virulence potential of CC10 strains and suggest lateral gene transfer and homologous recombination events that are unrelated to core phylogeny or source of isolation play a major role in CC10 diversity [36]. This also indicates that the underlying genetic factors responsible for the fitness and global spread of CC10 are unlikely to be related to our current understanding of virulence potential. This is an area for further investigation.

ExPEC-associated genes were among the most common genes detected in the collection. These included *fimH* (n=206, 83%)*, traT* (n=148, 60%)*, iss* (n=134, 54%) and iron acquisition-associated genes such as *irp2, fyuA, sitA, iutA* and *iucD* (n=88, 35%; n=92, 37%; n=56, 23%; n=23, 9%; n=44, 18%). *pap* operon pilus-assembly and *kpsMTII* capsular antigen genes were also detected. Carriage of ExPEC virulence factors has been previously reported in CC10 and may reflect the association of many of these genes with intestinal fitness [37], conversely it may

imply that innately fit and widespread lineages are simply more likely to acquire virulence genes. VAGs associated with enteric pathotypes were uncommon by comparison to ExPEC factors. Only 14 strains carried an *stx* variant characteristic of Shiga-toxigenic *E. coli* (STEC). One of these carried both *stx1B* and *stx2B.* Subtypes were not determined. These genes are typically acquired via bacteriophage transduction [38]. This observation highlights that abundant clonal groups such as CC10 may acquire virulence traits typical of both ExPEC and InPEC.

The average number of VAGs (n=6) in Australian porcine faecal strains was lower than the collection average (n=10). The most common ExPEC VAGs among porcine faecal strains were *fimH* (n=16/20, 8%), *traT* (n=15/20, 75%) *hek, irp2* and *fyuA* (all n=5/20, 25%). VAGs of intestinal pathotypes were not identified. Overall, the low abundance of VAGs suggests low virulence potential of these strains. In contrast, CC10 strains have been isolated from pigs with extraintestinal infections in China suggesting that porcine CC10 may acquire VAGs that cause ExPEC in the same host [32]. It is unknown whether these strains have developed virulence in the porcine gut via lateral gene transfer or if they are virulent strains introduced by humans to porcine production environments.

Resistance genes

We identified 73 ARGs in the total collection with a range of 0 to 19 ARGs per strain and an average of 5 per strain (Table S5.1). The average number of ARGs by origin was 7 for environmental, 6 for animals and humans and 4 for food-derived strains. Class 1 integrase gene *intI1* was present in 77 strains (31%) whilst class 2 integrase *intI2* was present in 29 strains (12%). Genes conferring resistance to older classes of antibiotics such as streptomycin (*strB/aph(6')-Id*; n=105, 42%, *strA/aph*(3`)*-Ib*; n=113, 46%), penicillin ($bla_{\text{TEM-1B}}$; n=93, 38%), aminoglycosides (*aadA1*; n=65, 26%), tetracycline (*tetA*; n=105, 42%) and sulphonamides (*sul2*; n=98, 40%) were most common in the collection. ESBL genes were not common, however they were represented by a variety of genes notably $bla_{\text{CTX-M-32}}$ (n=17, 7%), $bla_{\text{CTX-M-15}}$ (n=11, 4%), bla_{CMY} (n=13, 5%) and bla_{OXA1} (n=10, 4%).

ESBL carriage in this collection of CC10 is diverse but not particularly abundant. This is interesting given numerous reports in the literature pointing to CC10 as a common source of ESBL genes [21, 39, 40]. This once again demonstrates the current limitations data-mining and the need for more publicly available sequences to develop an accurate understanding of ESBL carriage in CC10 and *E. coli* collectively. This is particularly illustrated by the fact that ST10 was identified as the most common ESBL positive sequence type in Taiwanese river water [18]. The ability of CC10 to survive in river water has implications for its ability to disseminate and increases its exposure to diverse niches and ability to spread ESBL genes.

Colistin resistance gene *mcr1* was present in 5 sequences, four of which originated from wastewater, a known reservoir of antimicrobial resistance genes and mobile elements that transfer them [41]. Like ESBL carriage, *mcr* carriage in CC10 is commonly reported in the literature in contrast to the collection examined here [42-44]. A 2018 study by Garcia found that ST10 was the primary carrier of *mcr-4* and *mcr-5* variants in a collection of colistin-resistant porcine enterotoxigenic *E. coli* (ETEC) in Spain. This highlights the danger of exposing a successful lineage like CC10, within which the line between commensalism and pathogenicity is ill defined, to critically important antibiotics, as they are likely to acquire resistance rapidly.

The average number of ARGs for Australian porcine CC10 was 11, more than double that of the collection as a whole. This discrepancy is likely a result of the extensive history of antimicrobial use at the farm where the strains were isolated and selection of *intI1* positive strains as a proxy for multiple drug resistance [45]. A family of related integrons was abundant in this collection and responsible for a proportion of the extensive resistance [16]. The carriage of *sul3* in the Australian porcine strains was high (n=13/20; 65%) relative to the rest of the collection (n=11/228; 5%). This is likely a reflection of the lower carriage of class 1 integrons in the overall collection compared to the Australian isolates included in this study (all *intI1*$^{+}$) as *sul3* is associated with atypical class 1 integrons [16]. Carriage of *sul3* has an established association with *E. coli* from porcine sources [46]. ESBL and *mcr*

carriage was not observed in the Australian strains. It is difficult to compare our Australian porcine strains to the global strains with respect to antimicrobial resistance however it is clear that antimicrobial resistance depends on the exposure history of the strain as opposed to its physical or geographical source.

Plasmid replicons

We identified 21 plasmid Inc replicons in the collection. Strains carried between 0 and 7 replicons with an average of 2. IncF replicons dominated (FII; n=163 (66%), FIB; n=122 (49%), followed by IncI1 (n=66, 27%) and IncX1 (n=45, 18%). IncF and IncI1 plasmid families in particular are frequently implicated in the spread of antimicrobial resistance genes and are likely to play a role in the observed ARG carriage in this collection of CC10 [47]. The global expansion of ST131 ExPEC provides a cautionary tale in underestimating the potential for faecal commensal strains to pose a threat to human health. Like most ExPEC, ST131 is associated with human faeces and multiple IncF plasmids that carry extensive arrays of drug resistance genes. ST131, in conjunction with these plasmids, has expanded to become the predominant ESBL-producing clone in hospital and community-acquired ExPEC infections [7, 48]. It is conceivable that such a plasmid expansion event associated with a CC10 strain could yield a conjugative plasmid with serious antimicrobial resistance and virulence characteristics. As CC10 is a common faecal commensal of both animal and humans, can cause human ExPEC infections, inhabit diverse environmental niches and carry a wide variety of resistance-associated plasmids it is critical that this lineage is monitored in the context of human health.

Australian porcine CC10 also carried IncF replicons (FII, n=13 (65%), FIB; n= 13 (65%), IncI1 (n=7, 35%) and IncX1 (n=10, 50%) replicons. It is not known whether these plasmid types have been acquired on multiple occasions by CC10 or have remained stable in sub-lineages over long periods of time. IncHI2 was notably carried by 8 strains (40%), and was abundant in the larger collection of previously described porcine strains from Australia [16]. This is in contrast to the rest of the CC10 collection where IncHI2 was only present in 11 other strains. The IncHI2 replicons in the Australian porcine strains were all of the same pMLST type ST3, suggesting a localised plasmid acquisition event. This is supported by reports of

highly related IncHI2 plasmids with multiple resistance genes circulating in both *E. coli* and *Salmonella sp.* in the Asia Pacific region [49, 50]. This once again highlights the ability of CC10 and successful commensals to acquire diverse plasmids that carry drug resistance genes.

Mapping of plasmid and resistance gene carriage to the SNP tree did not appear to link phylogeny or origin with gene carriage (Fig. S5.3). Furthermore, MDS analysis of 211 unique plasmid replicon and resistance gene profiles did not cluster strains based on continent of isolation, sequence type or origin (Fig. 5.3(a-c)). Similar to the case with virulence genes, it appears that due to the widespread and diverse nature of CC10 accessory gene carriage is not associated with geography or source of isolation. However, it is possible that trends could be observed if sample geography or source was restricted. Similarly, trends might emerge if a larger, more comprehensively sampled collection were analysed. Further studies with controlled sampling and sufficient metadata are required to explore this.

Conclusions

This genomic analysis has demonstrated that CC10 is highly diverse with respect to the core genome as well as accessory elements. This diversity is illustrated by its host and geographic range and suggests a core set of fitness traits that are yet to be genetically defined and characterised. It is likely, as in the case of other global lineages of *E. coli,* that recombination and genomic islands in the chromosome play a role in its success. These attributes should be investigated in future work. Whilst this study is relatively large in scale, it is limited by our inability to control and balance sample sizes from different sources and geographic regions. As online databases of whole genome data continue to grow, the scale of these studies should be expanded to elucidate epidemiological features that cannot be determined currently. We suspect the apparent abundance of CC10 in human faeces indicates humans are the predominant intermediaries between strains found in other animals and environments. The dispersion of Australian porcine sequences throughout the phylogeny indicates they are derived from multiple lineages within CC10 that are well adapted to the porcine gut. The variety of VAGs

and ARGs suggests that these mobilised genetic traits are decoupled from phylogeny and depend instead on the history of each individual strain.

5.2.6 References

1. **Poolman JT, Wacker M**. Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *J Infect Dis* 2016;213(1):6-13.
2. O'Neill J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. London: Wellcome Trust, UK Government2016.
3. **Kaper JB, Nataro JP, Mobley HL**. Pathogenic Escherichia coli. *Nat Rev Microbiol* 2004;2(2):123-140.
4. **Fair RJ, Tor Y**. Antibiotics and bacterial resistance in the 21st century. *Perspect Medicin Chem* 2014;6:25-64.
5. **Roy Chowdhury P, Charles IG, Djordjevic SP**. A role for Tn6029 in the evolution of the complex antibiotic resistance gene loci in genomic island 3 in enteroaggregative hemorrhagic Escherichia coli O104:H4. *PLoS One* 2015;10(2):e0115781.
6. **Rohde H, Qin J, Cui Y, Li D, Loman NJ et al.** Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. *The New England journal of medicine* 2011;365(8):718-724.
7. **Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M et al.** Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A* 2014;111(15):5694-5699.
8. **Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM et al.** Genomic Microbial Epidemiology Is Needed to Comprehend the Global Problem of Antibiotic Resistance and to Improve Pathogen Diagnosis. *Front Microbiol* 2016;7:843.
9. **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.
10. **Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L et al.** Food reservoir for Escherichia coli causing urinary tract infections. *Emerg Infect Dis* 2010;16(1):88-95.
11. **Singer RS**. Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. *Front Microbiol* 2015;6:28.
12. **Nordstrom L, Liu CM, Price LB**. Foodborne urinary tract infections: a new paradigm for antimicrobial-resistant foodborne illness. *Front Microbiol* 2013;4:29.
13. **Dorado-Garcia A, Smid JH, van Pelt W, Bonten MJM, Fluit AC et al.** Molecular relatedness of ESBL/AmpC-producing Escherichia coli from humans, animals, food and the environment: a pooled analysis. *J Antimicrob Chemother* 2018;73(2):339-347.
14. **Manges AR, Harel J, Masson L, Edens TJ, Portt A et al.** Multilocus sequence typing and virulence gene profiles associated with Escherichia coli from human and animal sources. *Foodborne Pathog Dis* 2015;12(4):302-310.
15. **Manges AR, Johnson JR**. Food-borne origins of Escherichia coli causing extraintestinal infections. *Clin Infect Dis* 2012;55(5):712-719.
16. **Reid CJ, Wyrsch ER, Roy Chowdhury P, Zingali T, Liu M et al.** Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 2017;3(12).
17. **Liu X, Liu H, Li Y, Hao C**. High Prevalence of beta-lactamase and Plasmid-Mediated Quinolone Resistance Genes in Extended-Spectrum Cephalosporin-Resistant Escherichia coli from Dogs in Shaanxi, China. *Front Microbiol* 2016;7:1843.
18. **Chen PA, Hung CH, Huang PC, Chen JR, Huang IF et al.** Characteristics of CTX-M Extended-Spectrum beta-Lactamase-Producing Escherichia coli Strains Isolated from

Multiple Rivers in Southern Taiwan. *Applied and environmental microbiology* 2016;82(6):1889-1897.
19. **Varela AR, Manageiro V, Ferreira E, Guimaraes MA, da Costa PM et al.** Molecular evidence of the close relatedness of clinical, gull and wastewater isolates of quinolone-resistant Escherichia coli. *J Glob Antimicrob Resist* 2015;3(4):286-289.
20. **Tan C, Tang X, Zhang X, Ding Y, Zhao Z et al.** Serotypes and virulence genes of extraintestinal pathogenic Escherichia coli isolates from diseased pigs in China. *Veterinary journal (London, England : 1997)* 2012;192(3):483-488.
21. **Day MJ, Rodriguez I, van Essen-Zandbergen A, Dierikx C, Kadlec K et al.** Diversity of STs, plasmids and ESBL genes among Escherichia coli from humans, animals and food in Germany, the Netherlands and the UK. *J Antimicrob Chemother* 2016;71(5):1178-1182.
22. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA et al.** Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research* 2015;43(3):e15.
23. **Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T et al.** SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2016;2(4):e000056.
24. **Price MN, Dehal PS, Arkin AP**. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5(3):e9490.
25. **Letunic I, Bork P**. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* 2016;44(W1):W242-245.
26. **Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y et al.** ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;8(1):28-36.
27. **Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J et al.** ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3(10):e000131.
28. **Venables WN, Ripley BD**. *Modern Applied Statistics with S*: Springer Publishing Company, Incorporated; 2010.
29. **Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P et al.** *Vegan: Community Ecology Package. R Package Version. 2.0-10* 2013.
30. **Tsang AKL, Lee HH, Yiu SM, Lau SKP, Woo PCY**. Failure of phylogeny inferred from multilocus sequence typing to represent bacterial phylogeny. *Sci Rep* 2017;7(1):4536.
31. **Ronco T, Stegger M, Olsen RH, Sekse C, Nordstoga AB et al.** Spread of avian pathogenic Escherichia coli ST117 O78:H4 in Nordic broiler production. *BMC Genomics* 2017;18(1):13.
32. **Ding Y, Tang X, Lu P, Wu B, Xu Z et al.** Clonal analysis and virulent traits of pathogenic extraintestinal Escherichia coli isolates from swine in China. *BMC veterinary research* 2012;8:140.
33. **Schierack P, Walk N, Ewers C, Wilking H, Steinruck H et al.** ExPEC-typical virulence-associated genes correlate with successful colonization by intestinal E. coli in a small piglet group. *Environ Microbiol* 2008;10(7):1742-1751.

34. **Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E et al.** Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science (New York, NY)* 2005;307(5715):1618-1621.
35. **Han S**. *Environmental Impacts of China's Pork Industry*. Wilson Center; 2014.
36. **Hao W, Allen VG, Jamieson FB, Low DE, Alexander DC**. Phylogenetic incongruence in E. coli O104: understanding the evolutionary relationships of emerging pathogens in the face of homologous recombination. *PLoS One* 2012;7(4):e33971.
37. **Russell CW, Fleming BA, Jost CA, Tran A, Stenquist AT et al.** Context-Dependent Requirements for FimH and Other Canonical Virulence Factors in Gut Colonization by Extraintestinal Pathogenic Escherichia coli. *Infect Immun* 2018;86(3).
38. **Kruger A, Lucchesi PM**. Shiga toxins and stx phages: highly diverse entities. *Microbiology* 2015;161(Pt 3):451-462.
39. **Reuland EA, Overdevest IT, Al Naiemi N, Kalpoe JS, Rijnsburger MC et al.** High prevalence of ESBL-producing Enterobacteriaceae carriage in Dutch community patients with gastrointestinal complaints. *Clin Microbiol Infect* 2013;19(6):542-549.
40. **Wang J, Gibbons JF, McGrath K, Bai L, Li F et al.** Molecular characterization of blaESBL-producing Escherichia coli cultured from pig farms in Ireland. *J Antimicrob Chemother* 2016.
41. **Manaia CM, Rocha J, Scaccia N, Marano R, Radu E et al.** Antibiotic resistance in wastewater treatment plants: Tackling the black box. *Environment international* 2018;115:312-324.
42. **Garcia V, Garcia-Menino I, Mora A, Flament-Simon SC, Diaz-Jimenez D et al.** Co-occurrence of mcr-1, mcr-4 and mcr-5 genes in multidrug-resistant ST10 Enterotoxigenic and Shiga toxin-producing Escherichia coli in Spain (2006-2017). *Int J Antimicrob Agents* 2018.
43. **El Garch F, de Jong A, Bertrand X, Hocquet D, Sauget M**. mcr-1-like detection in commensal Escherichia coli and Salmonella spp. from food-producing animals at slaughter in Europe. *Vet Microbiol* 2018;213:42-46.
44. **Matamoros S, van Hattem JM, Arcilla MS, Willemse N, Melles DC et al.** Global phylogenetic analysis of Escherichia coli and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. *Sci Rep* 2017;7(1):15364.
45. **Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM et al.** Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *Isme j* 2015;9(6):1269-1279.
46. **Perreten V, Boerlin P**. A new sulfonamide resistance gene (sul3) in Escherichia coli is widespread in the pig population of Switzerland. *Antimicrob Agents Chemother* 2003;47(3):1169-1172.
47. **Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B et al.** Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* 2018.
48. **Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K et al.** Separate F-Type Plasmids Have Shaped the Evolution of the H30 Subclone of Escherichia coli Sequence Type 131. *mSphere* 2016;1(4).
49. **Chen W, Fang T, Zhou X, Zhang D, Shi X et al.** IncHI2 Plasmids Are Predominant in Antibiotic-Resistant Salmonella Isolates. *Front Microbiol* 2016;7:1566.

50. **Fang L, Li X, Li L, Li S, Liao X et al.** Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from E. coli isolates of food-producing animals. *Sci Rep* 2016;6:25312.

5.2.7 Data Bibliography

1. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al.** Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67, 2640–2644 (2012).
2. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS et al.** Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* 52, 1501–1510 (2014).
3. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58, 3895–3903 (2014).
4. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F.** Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 53, 2410–2426 (2015).

5.2.8 Tables and Figures

Source	Sequence Type								Total
	10	**34**	**43**	**44**	**48**	**167**	**215**	**218**	
Africa									**13**
Food	4		3						
Human	5		1						
Asia									**23**
Animal	2								
Environment	2	2	3		1				
Human	6			1		6			
Europe									**46**
Animal	5								
Food	10								
Human	27	1	1			2			
North America									**136**
Animal	37	2	1		4	1		1	
Environment	6		1						
Food	33	1	8		10		1	1	
Human	21	1	1	1	2	3			
Oceania									**30**
Animal	10				6			4	
Human	6	2			2				
									248
Total Source									
Animal	54	2	1		10	1		5	**73**
Environment	8	2	4		1				**15**
Food	47	1	11		10		1	1	**71**
Human	65	4	3	2	4	11			**89**
Total ST	**174**	**9**	**19**	**2**	**25**	**12**	**1**	**6**	**248**

Table 5.1

CC10 sequences used in this study.

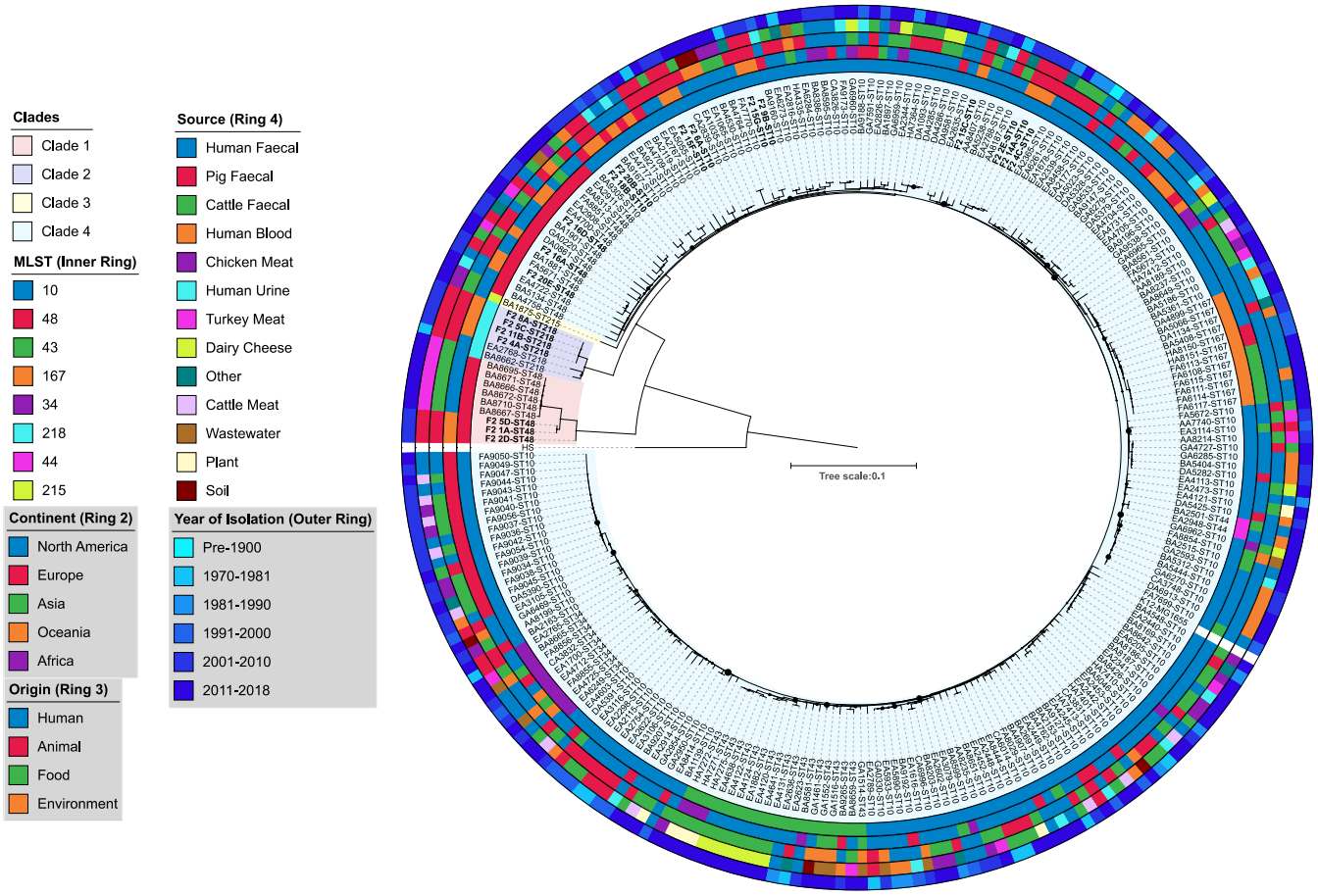


Figure 5.1

Maximum-likelihood phylogeny of 248 *E. coli* CC10 sequences. Australian porcine sequences are in bold. Small nodes indicate high confidence splits whilst larger nodes indicate lower confidence splits. Sequence types, continent of isolation, origin of isolation, source of isolation and year of isolation are annotated on coloured outer rings according to the legends. Clades are also coloured according to the legend. Metadata for reference strain K12-MG1655 and outgroup strain HS is omitted.

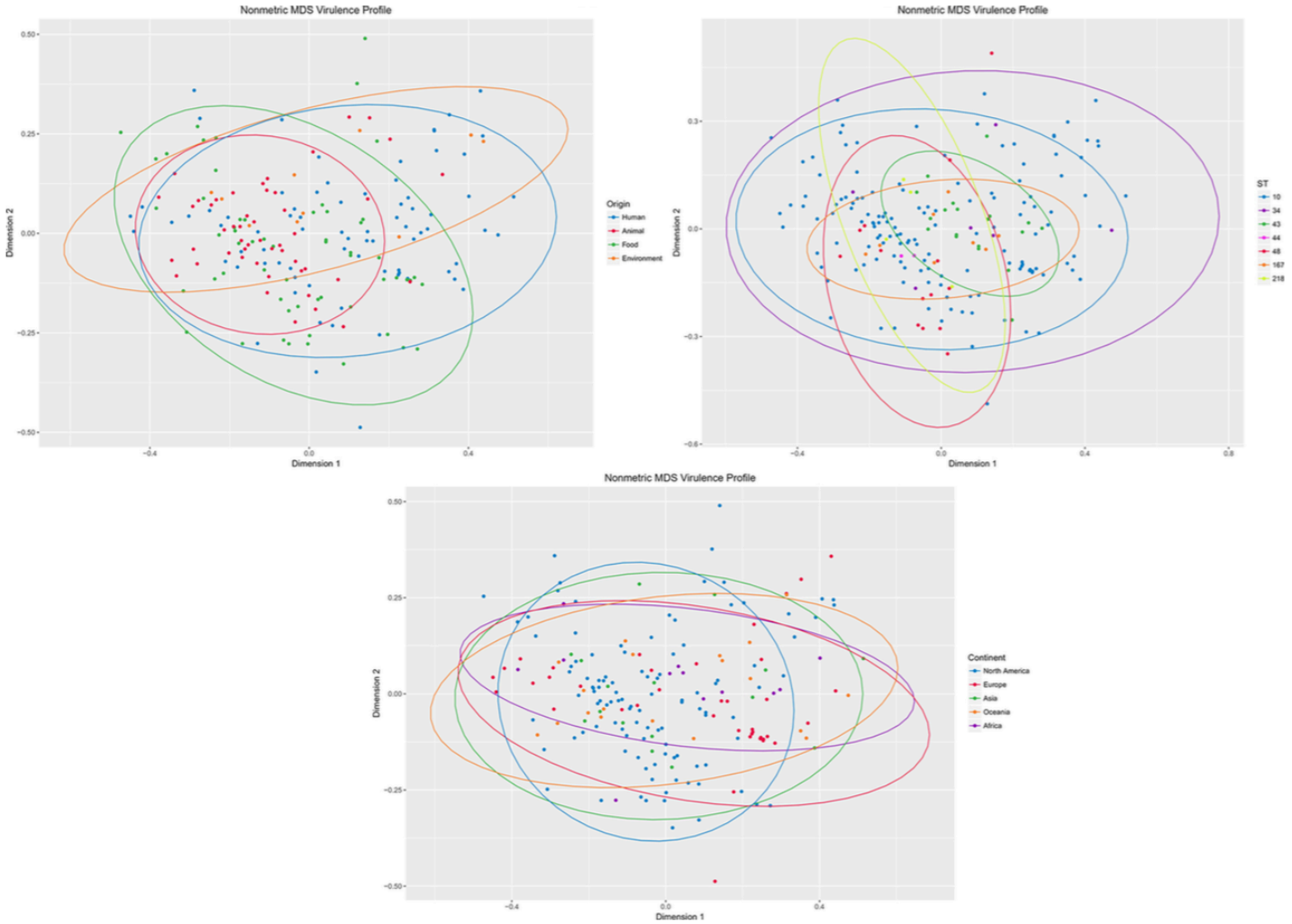


Figure 5.2

Non-metric MDS graph of 194 unique virulence gene profiles grouped by 95% confidence interval ellipses for a) continent of isolation b) origin of isolation and c) sequence type. N.B. Some characteristics contained too few data points for ellipses to be calculated.

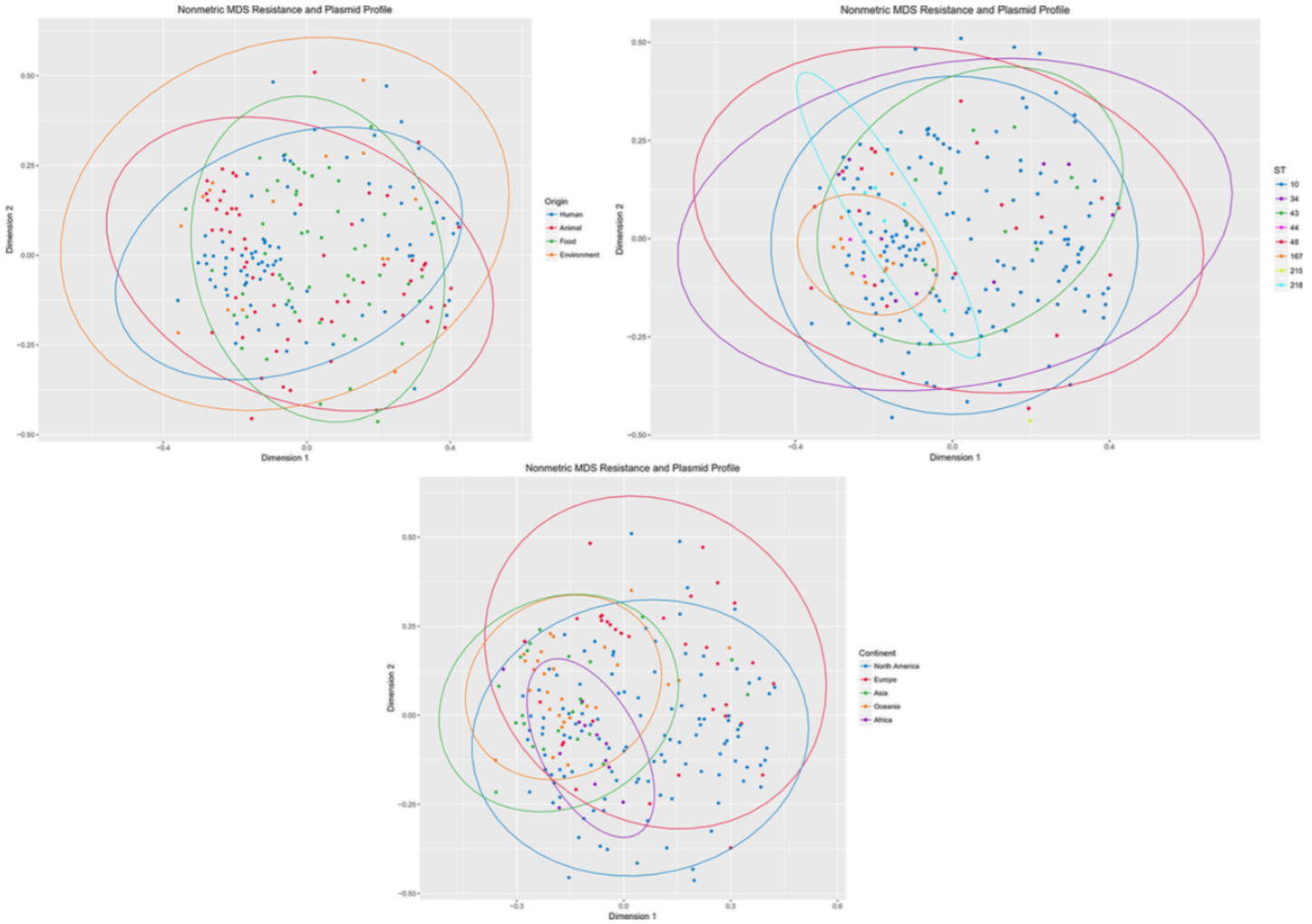


Figure 5.3

Non-metric MDS graph of 211 unique plasmid replicon and antimicrobial resistance gene profiles grouped by 95% confidence interval ellipses for a) continent of isolation b) origin of isolation and c) sequence type. N.B. Some characteristics contained too few data points for ellipses to be calculated.

Chapter 6: Comparative genomic analysis of two highly-related *E. coli* ST131-*H*22 strains from a healthy piglet and a human urinary tract infection

6.1 Declaration

This chapter is being prepared for publication but is yet to be submitted, therefore it is structured in the same manner as the previous chapters for consistency.

6.2 Comparative genomic analysis of two highly-related *E. coli* ST131-*H*22 strains from a healthy piglet and a human urinary tract infection

Having identified an ST131 strain in our porcine collection, we sought to compare it to sequences of human clinical strains of *E. coli* also being studied in our group. This preliminary analysis identified a highly similar ST131 strain isolated from catheter stream urine of a hospital patient with a urinary tract infection. It was apparent they carried identical integrons and potentially plasmids. We therefore utilised long read whole genome sequencing to compare them and assess the risk of the porcine ST131 strain to human health. Please view figures as PDFs attached to this thesis.

6.2.1 Abstract

The role of food production animals in contributing drug resistant ExPEC to human infections is incompletely elucidated. ST131 strains of the *H*30*Rx* subgroup, a pandemic uropathogen with conserved fluoroquinolone and cephalosporin resistance, are not frequently identified in animals. However, there are increasing reports of its precursor *H*22 within animals and associated meat products. Here we identified two highly related ST131-*H*22 strains from a healthy pig and a human infection in 2007 and 2009 respectively. We used both long and short genome sequencing to compare them to publicly available *H*22 genome sequences. Even within the context of *H*22 strains, a relatively clonal group, the two strains in question were highly related, separated by only 47 core SNPs. Furthermore, they were closely related to a faecal strain isolated from a healthy human in NSW. Both strains carried highly similar IncHI2-ST3 multidrug resistance plasmids with differences in the hospital strain mainly arising due to IS mediated insertions and

rearrangements. Near identical ColV plasmids were also present in both strains, further supporting their shared evolutionary history. It is possible these strains represent a case of zoonotic transfer of a drug resistant pathogen from pigs to humans though more data is required to confirm that this is the case.

6.2.2 Data Summary

1. Short and long reads for both F2_14D and 2009_36 have been uploaded to SRA under BioProject accession number PRJNA508590.

6.2.3 Introduction

Multi-drug resistant (MDR) extra-intestinal pathogenic *E. coli* (ExPEC) that cause urinary tract infections (UTI), pyelonephritis and urosepsis represent a significant healthcare burden worldwide [1]. Whilst a diversity of *E. coli* clones representing various multi-locus sequence types may cause extra-intestinal infections[2], much recent study has focused on the globally disseminated ST131 *H*30*Rx* sub-lineage of ST131. ST131 *H*30*Rx* causes a significant proportion of hospital and community-acquired urine and blood-related infections and is resistant to last-line clinical antibiotics such as fluoroquinolones and cephaosporins[3]. Cephalosporin resistance, conferred by CTX-M-type extended-spectrum beta-lactamase expression, in particular contributes to its predominance in ExPEC literature. ST131 as a clonal group exhibits subpopulation clonal structure that correlates with carriage of different alleles fimbrial adhesin gene *fimH*, which facilitates adherence to uroepithelium, bladder cell invasion and establishment of intracellular bacterial communities[4]. Three major *fimH* alleles, *fimH*30*,* *H*22 and *H*41 dominate ST131 phylogeny with the latter two being basal to *H*30[5, 6]. *H*22 strains, which gave rise to *H*30*Rx*, are underrepresented in literature and sequence databases, probably due to the natural bias towards clinical and multi-drug resistant strains as these have the most immediate impact on human health. However, *H*22 also cause serious extra-intestinal infections[7] and ignorance of these ancestral, yet extant lineages, obscures the factors that established ST131 before the *H*30 lineage acquired ESBL genes. Increasing evidence indicates that ST131, and *H*22 in particular, are successful commensals [8]. Recent evidence showed that *H*22 strains isolated from healthy humans might have an advantage in the human gut due to better biofilm formation and use of gluconate as a carbon

source[9]. Furthermore, the global distribution of *H*22 and carriage of plasmids involved in intestinal fitness supports the contention that widespread faecal carriage was the initial reason for their success[10].

In addition to human carriage, *H*22 is also reported in wild seals and poultry meat, and was recently identified carrying colistin resistance determinant *mcr1* in swine from Spain[10-12]. A body of literature suggests intensive animal production in particular plays a role in the selection and emergence of drug resistant ExPEC however many of these studies employ restrictive selection criteria and are limited to comparison of known antimicrobial resistance genes, virulence-associated genes and PFGE profiling[13-15]. The presence of ST131 sub-clones in animals is probably attributable to humans in the first instance, however this does not exclude animals from playing a role in its ongoing evolution and dissemination as seen in the Spanish study where colistin resistance has clearly emerged in response to the use of this antimicrobial in swine production[12]. Regardless of origin, the presence of a human pathogen in animals is a concern as resistance, virulence and fitness traits encoded on mobile DNA that circulate in animal production may drive the evolution of even more severe human pathogens. In order to truly understand the evolution, reservoirs and dissemination of *H*22, far larger collections of systematically sampled, temporally and geographically related populations of strains from humans, animals and the environment need to be considered using a whole genome sequencing approach. SNP-based core genome phylogenies and long read sequencing characterisation of resistance and virulence plasmids provide superior resolution to MLST and screening of individual genes alone. In the absence of these large datasets, insights can still be gathered with the aforementioned methods. By understanding *H*22 and other established pre-pandemic lineages we may be able to predict future pathogen expansion events.

Here we characterised and compared two highly related *H*22 strains from disparate sources. F2_14D was isolated in 2007 from a healthy piglet at a rural production facility and 2009_36 was isolated in 2009 from a human urinary tract infection at a suburban hospital over 250km away. We aimed to compare their

ARGs, VAGs and plasmid types to a global collection of *H*22 strains and perform long-read sequencing to characterise the plasmids they carried.

6.2.4 Methods

Strains and sequences used in this study

Escherichia coli strain F2_14D was isolated from a weaned piglet by faecal swab in 2007 from an intensive production system with a history of extensive neomycin use for treatment of enterotoxigenic *E. coli* (ETEC) outbreaks. More information on the collection is previously reported [16]. 2009_36 was isolated from catheter stream urine of a patient with a urinary tract infection at the Sydney Adventist Hospital in Sydney in 2009. Enterobase (http://enterobase.warwick.ac.uk/species/index/ecoli) was queried for ST131 and strains designated *fimH*22 or clade B were downloaded from SRA using parallel-fastq-dump v0.6.3 (https://github.com/rvalieris/parallel-fastq-dump). We also included 11.3-R3, a human faecal *H*22 strain from Australia[17]. Both the raw Illumina reads and complete genome of JJ1897, an ST131 *H*22 isolated from a human infection were downloaded for gene screening and use as a reference genome respectively. These 48 strains and their accession numbers are available in Table S6.1.

DNA Isolation, Sequencing and Assembly

DNA from both strains was isolated, quantified and sequenced on an Illumina HiSeq 2500 v4 sequencer as previously described[16]. DNA was also isolated from both strains for long read sequencing by phenol-chloroform extraction (full method available in File S6.1, See Appendix) and quantified by Qubit dsDNA HS assay (Thermo Fisher Scientific) as previously described[16]. Suitable molecular weight for long read sequencing was confirmed on a 0.8% agarose gel, run at 10V for 16 hours and purity confirmed by Nanodrop (Thermo Fisher Scientific). Libraries were prepared for long-read sequencing using the Oxford Nanopore Technologies (ONT) 1D ligation sequencing kit (SQK-LSK108) with the native barcoding expansion kit (EXP-NBD103). Several modifications were made to the ONT protocol to maximise read length and throughput, including those described by Wick et al. (2017). In addition, we used 7.5 µg of starting DNA from each isolate

and performed the DNA purifications steps using SPRIselect beads (Beckman Coulter). Resuspension of SPRIselect beads was carried out at higher than usual temperatures (50°C after end repair and 37°C after adapter ligation) to promote efficient elution of the DNA into solution. The final library containing 4.4 µg DNA was loaded onto an ONT MinION instrument with a FLO-MIN106 (R9.4) flow cell and run for 48 h as per manufacturer's instructions. Raw fast5 files were base-called with Albacore v2.3.3 (ONT) and de-multiplexed with Porechop v0.2.3 (https://github.com/rrwick/Porechop). Reads were then subsampled to 500Mbp with Filtlong v0.2.0 with default settings and a minimum read length of 2000bp. Unicycler v0.4.6 [18] hybrid assembly was performed with default settings using both Illumina raw reads and Oxford Nanopore subsampled reads. Contigs less than 1000bp were excluded from the final assembly. Assemblies graphs viewed in Bandage [19] revealed the chromosome of 2009_36 was fragmented however two circular contigs of 278,665bp and 143,671bp were assembled. F2_14D consisted of a single contig circular chromosome and a circular contig of 274,883bp. Three further contigs with a continuous graph path were joined and exported from Bandage in fasta format, forming the third circular replicon 139,372bp in length. All sequence reads, both short and long, used in this study have been uploaded to SRA under BioProject PRJNA508590. Individual Accession numbers are available in Table S6.1.

Phylogenetic Analysis

In order to generate a global ST131-*H*22 core SNP phylogeny, Illumina raw reads were aligned to the complete genome of *H*22 strain JJ1897 (gb|CP013837.1) using Snippy v4.1.0 (https://github.com/tseemann/snippy). The completed genome of ST131-*H*30 strain EC958 (gb|NZ_HG941718.1) was also aligned as an out-group strain to root the phylogeny. snippy-core was then used to generate an alignment of all strains. The snippy-core full alignment was recombination filtered with Gubbins, and SNPs identified with snp-sites v2.4.0 resulting in an alignment of 2292 core variable sites[20]. FastTree2 v 2.1.10 was then used with default settings to produce maximum-likelihood phylogenetic trees for both alignments using a generalised time-reversible (GTR) nucleotide substitution model[21]. The tree was visualised with strain metadata in iTOL v4.2.3[22]. The tree was also

visualised alongside a gene-screening heatmap in R v3.3.1 with ggtree v3.6[23]. Pairwise SNPs between all strains were extracted from the core alignment using snp-dists v0.6 (https://github.com/tseemann/snp-dists). The Gubbins filtered alignment (gubbinsfullcore.aln), final cleaned alignment (cleanfullcore.aln), tree (cleanfullcore.tree) and the pairwise SNP table (pairwise_snps.csv) are available at https://github.com/CJREID/ST131-H22_supporting_data.

Gene screening

All *H*22 sequences were screened for ARGs, VAGs, plasmid replicons, O and H antigen genes and pMLST alleles using ARIBA v 2.10.1[24] and ResFinder, PlasmidFinder, VirulenceFinder and SerotypeFinder databases available from the Center for Genomic Epidemiology (http://www.genomicepidemiology.org/) [25-28]. A custom database of additional genes not present in the aforementioned databases was also used[29].

Plasmid annotation and visualisation

Circular plasmid contigs resulting from the Unicycler assembly were annotated with RASTtk via the Pathosystems Resource Integration Centre's (PATRIC) online Bacterial Bioinformatics Resource Centre (https://patricbrc.org/). This automated annotation was then imported in GenBank format into SnapGene 4.1.9 (GSL Biotech) and manual annotation was performed with BLASTn (https://blast.ncbi.nlm.nih.gov/) and the previously mentioned gene databases. The complete annotated plasmid sequences of pF2_14D_HI2, p F2_14D_F, p2009_36_HI2 and p2009_36_F have been uploaded to Genbank though accession numbers have not been confirmed at the time of this thesis submission. NCBI BLASTn was also used to select complete plasmids for comparison to our sequence. For F plasmids, mobile elements and transposons were removed from the backbone of p2009_36_F and the backbone sequence was used to query GenBank. Complete plasmid sequences with 100% coverage and ≥97% identity were downloaded for comparison (n=13). For HI2 plasmids, the *smr0018* and *smr0199* alleles were used as a query to identify and download all publicly available IncHI2-ST3 plasmids (n= 41) (accessed 25/10/18). Details of downloaded plasmids are available in Table S6.2-S6.3. BLAST Ring Image

Generator (BRIG) v0.95 was used with default settings to compare our plasmid sequences to the publicly available ones[30]. These figures were then combined with the SnapGene generated plasmid maps to visualise similarities and differences.

6.2.5 Results and Discussion

F2_14D and 2009_36 in the context of ST131-H22 global phylogeny

Routine preliminary analysis (data not shown) undertaken on our hospital and porcine origin *E. coli* sequences indicated that two ST131 strains from these collections were closely related and had similar accessory genes. These strains were 2009_36, isolated from a human catheter stream infection in 2009 and F2_14D, isolated from a faecal swab of a healthy piglet in 2007.

In order to determine how similar these two strains were at a core genome level we performed SNP-based alignment with snippy, using the closed genome of ST131-H22 JJ1897 as a reference and 47 publicly available *H*22 strains. The closed genome of ST131-*H*30 strain EC958 was used to root the maximum-likelihood tree. The core genome alignment used to build the tree consisted of 2,292 conserved variable sites. (Fig. 6.1)

Publicly available strains were from human (n= 42), avian (n=3), food (n=1) and undetermined (ND, n=2) sources, spanning 1967-2012, four continents and seven countries (Fig. 6.1). Human-sourced strains originated from blood (n=12), urine (n=16) and respiratory (n=1) infections. A single Australian faecal isolate from a healthy human was also included (Table S6.1). The lack of human faecal-derived strains was surprising given this is typically the immediate source of extra-intestinal infections and human faecal carriage of *H*22 is previously reported [9, 31, 32].

Aside from JJ1897, which was present on its own clade, the *H*22 strains split into three well-supported clades designated A, B and C (Fig. 6.1, Fig. 6.2). The structure of our tree is similar to the phylogeny observed by Liu *et al* [10]. The tree did not clearly stratify strains based on origin, country or year of isolation. However a few

clusters of geographically related strains were apparent. 2009_36 and F2_14D were present in clade A and separated by 47 SNPs. Human faecal strain 11.3-R3 from 2010 was also closely related to these two strains separated by 60 and 51 SNPs from 2009_36 and F2_14D respectively. Whilst more examples of this lineage from animals and humans are required to understand potential reservoirs, it is remarkable that three strains from randomly sampled collections are so highly related. This strongly suggests this lineage is widespread and abundant, likely within the human gut. The fact the porcine strain was the only ST131 among 68 isolates from 21 pigs suggests its presence in swine is attributable to humans, however further studies of porcine commensal *E. coli* are required to explore this hypothesis[16]. Two further sequences from human blood infections in the UK (2010 and 2012) were present on the same sub-clade, separated by less than 90 SNPs from our strains, indicating this lineage of *H*22 is also globally disseminated and may have been introduced to Australia from overseas.

ST131-H22 virulence profiles reflect phylogeny

In order to determine the repertoire of accessory elements in *H*22 strains and examine any patterns in the context of the phylogeny we screened for virulence, resistance, plasmid-associated and common mobile element genes and mapped them against the core genome phylogeny (Fig. 6.2). Uropathogenic specific protein gene *usp* and ferric yersiniabactin uptake gene *fyuA* were present in all strains. Genes involved in iron acquisition *irp2* (n=48, 96%)*, iucD* (n=46, 92%)*, iutA* (n=45, 90%) and *sitA* (n=46, 92%) were abundant, as were immune evasion/protectin genes *kpsMTII* (n=48, 96%)*, traT* (n=47, 94%), *iss* (n=49, 98%) and invasion of brain endothelium gene *ibeA* (n=48, 96%)[33-40]. Generally, each clade appeared to have distinguishable virulence gene profiles with clades A and B being more similar than clade C.

ColV carriage is associated with multiple IncF plasmids in clades A and B

Clades A and B were notable for their carriage of ColV genes *cvaABC* and *cvi,* which were absent from clade C. ColV plasmids are strongly associated with avian pathogenic *E. coli* (APEC) where they provide an advantage in the gut and virulence in extra-intestinal sites they infect. Furthermore they have been

identified in human commensals carrying ARGs and human urine and blood infections, indicating they play a similar role dual here[10, 31, 41-43]. Fifteen strains, including 2009_36 and F2_14D carried all four of these genes whilst three others carried at least one of them. Mapping of ColV-associated gene carriage to the SNP phylogeny shows that two clades of ST131-*H*22 carry ColV plasmids (Fig. 6.2). A total of eight F-plasmid types corresponded to carriage of ColV genes in the collection with F18:B1 (n=5) and F2:B1 (n=5) being the most common. These results are similar to those of Liu *et al* and our additional pMLST data supports the repetitive acquisition of different ColV plasmids by these lineages of *H*22[10]. The carriage of ColV plasmids in these lineages suggests they have been exposed and adapted to niches where these plasmids are common such as poultry production systems and the human gut[10, 41].

Antimicrobial resistance in H22

Twenty-three strains carried no ARGs whilst the remaining strains carried between 1 and 11. ARGs conferring resistance to aminoglycosides (*aadA1,* n=9, 18%), penicillin (*bla*$_{\text{TEM-1B}}$, n=9, 18%), streptomycin (*strA,* n=9, 18%; *strB,* n=8, 16%), sulphonamides (*sul1,* n=8, 16%; *sul2*, n=7, 14%) and tetracycline (*tetA,* n=10, 20%) were most common. These genes are frequently encountered in commensal and pathogenic *E. coli* that inhabit the gut of swine in Australia[16, 44, 45]. Furthermore, IS*26*-flanked transposons carrying these genes, such as Tn*6029* and Tn*6026* are frequently encountered in food animals and humans in Australia on diverse plasmid backbones[31, 46-50]. It is notable that these transposons are also present on virulence plasmids including ColV and IncI plasmids in human pathogenic *E. coli.* Class 1 integrase gene *intI1* was present in ten strains (20%), all of which carried multiple ARGs consistent with the association between class 1 integrons and multidrug resistance[51]. Strains carrying ARGs typically carried a non-F type plasmid replicon. This indicates that the history of antimicrobial exposure for each strain remains important within *H*22 and they continue to acquire diverse plasmid types. Despite the apparent susceptibility or carriage of 'older' ARGs in *H*22, ESBL genes were present in the collection. Two strains carried *bla*$_{\text{OXA-2}}$, whilst *bla*$_{\text{CMY-2}}$, *bla*$_{\text{CTX-M-1}}$, *bla*$_{\text{OXA-1}}$, and *bla*$_{\text{SHV-2}}$ were each present in single strains. This highlights the ease with which *H*22 strains might acquire serious drug

resistance characteristics; especially given they are isolated from human blood in hospitals where antimicrobial selection pressure abounds.

Plasmids in 2009_36 and F2_14D

We used long-read sequencing to determine the similarity between the apparently identical plasmids carried by 2009_36 and F2_14D. This confirmed each strain carried an F2:B1 ColV plasmid and an IncHI2-ST3 plasmid. These plasmids were designated p2009_36_F, pF2_14D_F, p2009_36_HI2 and pF2_14D_HI2.

p2009_36_HI2 and pF2_14D_HI2 are large IncHI2-ST3 antimicrobial and metal resistance plasmids

p2009_36_HI2 and pF2_14D_HI2 were 278,665bp and 274,883bp long respectively and shared 99% sequence identity across 98% of their length. Most of the differences in gene content were due to acquisition of insertion elements and truncations caused by them. Relative to pF2_14D_HI2, p2009_36_HI2 carried five additional copies of IS*1203*-like elements, 6 additional copies of IS*26* and a copy of IS1294*.* These IS elements appear responsible for rearrangements in the backbone of p2009_36_HI2. Due to multiple insertions and rearrangements it is difficult to accurately determine the process of evolution that led to the structure of p2009_36_HI2, however it appears that one or more insertions of IS*26* has led to a large inversion relative to pF2_14D_HI2 (Fig. 6.3, Fig. 6.4)[52]. IS acquisition and rearrangements in p2009_36_HI2 may reflect its exposure to hospital environments where antimicrobial use can induce the SOS response and up-regulate replication and transposition of IS elements[53]. Aside from these differences, both plasmids were remarkably similar. Both carried two identical class 1 integron structures. The first of these was an IS*26* flanked *sul3*-associated integron lacking direct repeats, suggesting that an IS*26*-related homologous recombination event mediated its insertion. It consists of Δ*intI1* truncated by a copy of IS*26,* followed by a cassette array *estX-psp-aadA2-cmlA-aadA1.* The 3`-*CS* is a variant of that previously described in pCERC3[42]. IS*1203* is inserted within *qacH* followed by putative transposase *tnp440* and a downstream module IS*26-ΔmefB-orfB-orfA-sul3*-IS*26.* Only 111bp of macrolide efflux gene *mefB* remains and the module is inverted relative to the structure in pCERC3 such that Δ*mefB* is

proximal to the integron. This deletion signature was common in the collection that F2_14D originates from and all other strains carrying it also carried an IncHI2 replicon[16]. This strongly suggests that plasmids related to pF2_14D_HI2 are responsible for the carriage of this integron in that collection. Related *sul3* integrons have also been reported in IncF and IncI1plasmids, however this signature is only reported in Australian pigs to date[16, 42, 54, 55].

The second integron is a *sul1* integron located downstream of a Tn*1721*-like structure carrying *tetA.* It carries a cassette array *dfrA12-orfF-aadA2* followed by 3`-*CS* in which *tniA* is truncated by IS*26.* This integron and relatives are similarly widely reported[56]. The *pco/sil* operon, conferring copper and silver resistance, associated with a suite of Tn*7*-like transposase genes was present in both plasmids as well as the *ter* gene cluster conferring tellurite resistance[57, 58].

p2009_36_HI2 and pF2_14D_HI2 are closely related to HI2-ST3 resistance plasmids circulating in the Asia-Pacific region

Using the pMLST alleles as a query, we downloaded all 41 complete HI2-ST3 plasmid sequences for comparison with pF2_14D_HI2 and p2009_36_HI2. All bar two sequences originated from China and Hong Kong, reflecting numerous reports indicating they are endemic to the Asia-Pacific region[57, 58]. The two non-Chinese sequences (pIncHI2-MU3 and pSTM-275) were both from Australian pigs, one originating from *E. coli* and the other from *S. enterica* species respectively. pSTM-275 was nearly identical to pF2_14D_HI2 and carried an identical *sul3* type integron and a variant of the *sul1* integron. They differed only in the presence of a number of additional IS elements in pF2_14D_HI2. pIncHI2-MU3 by contrast did not carry either of the two integrons, or the *neoR* gene present in F2_14D_HI2. The Australian sequences were generally more similar to each other than to non-Australian strains indicating a potentially Australia-specific lineage of HI2 plasmids. Six Chinese sequences also originated from pigs. This supports a local plasmid acquisition event by the ancestral strain, possibly within pigs. The presence of HI2-ST3 in human faeces, infections, swine, poultry, sewage, lettuce and even giant panda suggests these plasmids provide a fitness advantage to gram-negative bacteria across a wide variety of niches. Furthermore, carriage in multiple

E. coli lineages*, K. pneumoniae, R. ornitholytica* and *S. enterica spp.* indicates they are frequently transferred within and between species relevant to human health. Altogether this data shows there is a need for understanding local patterns in AMR plasmid carriage to better manage the risks associated with their acquisition by pathogens such as ST131.

The HI2-ST3 backbone containing conjugative transfer genes, a tellurite resistance operon and numerous hypothetical proteins appeared to be highly conserved across all sequences. Differences between plasmids were primarily due to insertion sequences and drug resistance regions. (Fig. 6.3, Fig. 6.4). The presence of the *pco/sil* resistance operon in tandem with ARGs in all three Australian porcine strains is concerning. The use of copper as an in-feed additive for production swine and use of antimicrobials for disease outbreaks provides a dual selection pressure for these plasmids in swine and explains their carriage in porcine populations of *E. coli* [59, 60]. The major issue is that even in the absence of antimicrobial use, these plasmids and their ARGs will continue to be conserved and pass into environments where they may acquire further ARGs. This is illustrated by the Chinese sewage plasmid pMCR_WCHEC050613, which carried the *pco/sil* operon, variants of the integrons described in our plasmids and mobile colistin resistance gene *mcr-1*. The contamination of wastewater with antimicrobials and heavy metals is likely to drive this sort of rapid evolution that constitutes a threat to human health[61, 62]. The presence of an IS-flanked neomycin resistance gene *neoR* in our two plasmids and another Australian porcine sequence pSTM6-275, and absence from all other sequences, is intriguing and could be related to the common use of neomycin in Australian pigs[63]. Further work is required to determine the full range and transmission of HI2-ST3 in the Asia-Pacific region and these plasmids should be monitored in the context of human health.

2009_36 and F2_14D carry closely related ColV F2:B1 virulence plasmids associated with pathogenicity

p2009_36_F and pF2_14D_F are ColV virulence plasmids with identical F2:B1 replicons, 143,671bp and 139,372bp in length respectively. pF2_14D_F carried a

copy of IS*Ec23* that was absent from p2009_36_F whilst p2009_36_F carried an IS*1203*-like element absent from pF2_14D_F. Both plasmids carried Tn*2,* inserted within a Tn*1721*-derivative transposon however the *tet* genes and part of *tnpA-1721* were deleted in pF2_14D_F. Otherwise, these two plasmids were identical in structure (Fig. 6.5, Fig. 6.6). They carried *cvaABC* and *cvi* colicins, as well as numerous VAGs including *iroBCDEN* salmochelin operon, *etsABC* type 1 secretion system, *iucABCD* and *iutA* aerobactin operon, *hlyF, ompT* and *iss*[41, 64, 65].

We queried GenBank with the shared backbone of the two plasmids, free of IS elements and Tn*1721* derivatives, and downloaded complete plasmids with ≥97% sequence homology across their length for comparison. Sequences originated from human faecal, clinical and poultry sources, reflecting previously reported reservoirs where ColV plasmids may confer a selective advantage. The geographical distribution included Australia, Europe and North America however Asia was not represented. Our plasmids were most similar to the pCERC plasmids from human faecal *E. coli* described by Moran and Hall[31, 42]. pCERC5 was the most similar, carrying the same backbone and Tn*1721*-derivative as p2009_36_F, the only difference being an additional insertion of an IS*1203*-like element in p2009_36_F (Fig. 6.5., Fig. 6.6). Remarkably, pCERC5 was isolated from 11.3-R3, the closest relative of our two strains on the SNP tree. In contrast to the carriage of large IncHI2 drug resistance plasmids, the two ARGs in this strain were attributable to pCERC5, though it carried an additional IncN replicon (Table S6.1).

Transfer of strains and plasmids

The nearly identical core genomes and plasmid carriage of F2_14D, 2009_36 and 11.3-R3 are intriguing, and raise more questions than they answer. Firstly, the conserved carriage of F2:B1 ColV plasmids in a limited evolutionary background indicates that a common ancestor of these strains acquired this plasmid. Whilst not the only possibility, we believe there are several reasons that the case of F2_14D and 2009_36 indicates a transfer event of their common ancestor from pigs to humans. Firstly, most HI2-ST3 plasmids don't carry the *pco/sil* operon. p2009_36_HI2 was the only human clinical sequence examined that possessed this locus. The others were from Australian pigs and Chinese lettuce and sewage.

Heavy metal resistance is likely to be enriched in all three of these settings by the selective pressure of copper in pig feed and animal manure that is used to fertilise crops, as well as heavy metal contamination of sewage[60, 62, 66]. Secondly, the *sul3* integron with identical *mefB* deletion is a genetic signature that, to date is only reported in Australian pigs, suggesting it evolved and disseminated from this niche [16, 54]. Thirdly, the extensive IS-mediated rearrangement in p2009_36_HI2 indicates it has evolved from pF2_14D_HI2 or a very similar plasmid, possibly in a hospital environment[53]. It is probable that the ancestor of both strains was introduced to pigs via humans, however the presence of the HI2-ST3 plasmid and its specific gene carriage suggests the subsequent event is backflow from pigs to humans. We concede that far more long-read sequencing of closely related strains is required to confirm this hypothesis, however the observations and implications emerging from such a small dataset are quite surprising. Whilst animal production may not be the true origin of human-adapted pathogens such as ST131, it may nonetheless play an ongoing role in their evolution.

6.2.6 References

1. **Russo T, Johnson J**. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes and Infection* 2003;5(5):449-456.
2. **Manges AR, Tabor H, Tellis P, Vincent C, Tellier PP**. Endemic and epidemic lineages of Escherichia coli that cause urinary tract infections. *Emerg Infect Dis* 2008;14(10):1575-1583.
3. **Nicolas-Chanoine MH, Bertrand X, Madec JY**. Escherichia coli ST131, an intriguing clonal group. *Clinical microbiology reviews* 2014;27(3):543-574.
4. **Kaushal D, Totsika M, Beatson SA, Sarkar S, Phan M-D et al.** Insights into a Multidrug Resistant Escherichia coli Pathogen of the Globally Disseminated ST131 Lineage: Genome Analysis and Virulence Mechanisms. *PLoS ONE* 2011;6(10):e26578.
5. **Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M et al.** Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A* 2014;111(15):5694-5699.
6. **Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE et al.** Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. *MBio* 2016;7(2):e02162.
7. **Morales-Barroso I, Lopez-Cerero L, Molina J, Bellido M, Navarro MD et al.** Bacteraemia due to non-ESBL-producing Escherichia coli O25b:H4 sequence type 131: insights into risk factors, clinical features and outcomes. *Int J Antimicrob Agents* 2017;49(4):498-502.
8. **Leflon-Guibout V, Blanco J, Amaqdouf K, Mora A, Guize L et al.** Absence of CTX-M Enzymes but High Prevalence of Clones, Including Clone ST131, among Fecal Escherichia coli Isolates from Healthy Subjects Living in the Area of Paris, France. *Journal of clinical microbiology* 2008;46(12):3900-3905.
9. **Nicolas-Chanoine MH, Petitjean M, Mora A, Mayer N, Lavigne JP et al.** The ST131 Escherichia coli H22 subclone from human intestinal microbiota: Comparison of genomic and phenotypic traits with those of the globally successful H30 subclone. *BMC microbiology* 2017;17(1):71.
10. **Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K et al.** Escherichia coli ST131-H22 as a Foodborne Uropathogen. *MBio* 2018;9(4).
11. **Mora A, García-Peña FJ, Alonso MP, Pedraza-Diaz S, Ortega-Mora LM et al.** Impact of human-associated Escherichia coli clonal groups in Antarctic pinnipeds: presence of ST73, ST95, ST141 and ST131. *Scientific Reports* 2018;8(1).
12. **Garcia-Menino I, Garcia V, Mora A, Diaz-Jimenez D, Flament-Simon SC et al.** Swine Enteric Colibacillosis in Spain: Pathogenic Potential of mcr-1 ST10 and ST131 E. coli Isolates. *Front Microbiol* 2018;9:2659.
13. **Manges AR, Johnson JR**. Food-borne origins of Escherichia coli causing extraintestinal infections. *Clin Infect Dis* 2012;55(5):712-719.
14. **Singer RS**. Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. *Front Microbiol* 2015;6:28.
15. **Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L et al.** Food reservoir for Escherichia coli causing urinary tract infections. *Emerg Infect Dis* 2010;16(1):88-95.
16. **Reid CJ, Wyrsch ER, Roy Chowdhury P, Zingali T, Liu M et al.** Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 2017;3(12).

17. **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.
18. **Wick RR, Judd LM, Gorrie CL, Holt KE**. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13(6):e1005595.
19. **Wick RR, Schultz MB, Zobel J, Holt KE**. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015;31(20):3350-3352.
20. **Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T et al.** SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial genomics* 2016;2(4):-.
21. **Price MN, Dehal PS, Arkin AP**. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5(3):e9490.
22. **Letunic I, Bork P**. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* 2016;44(W1):W242-245.
23. **Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y et al.** ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;8(1):28-36.
24. **Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J et al.** ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3(10):e000131.
25. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O et al.** In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58(7):3895-3903.
26. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS et al.** Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol* 2014;52(5):1501-1510.
27. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F**. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2015;53(8):2410-2426.
28. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S et al.** Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67(11):2640-2644.
29. **Reid CJ, DeMaere MZ, Djordjevic SP**. Australian porcine clonal complex 10 (CC10) Escherichia coli belong to multiple sublineages of a highly diverse global CC10 phylogeny. *Microb Genom* 2018.
30. **Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA**. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, journal article 2011;12(1):402.
31. **Moran RA, Hall RM**. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. *Microbial drug resistance (Larchmont, NY)* 2017.
32. **Rodrigues C, Machado E, Fernandes S, Peixe L, Novais A**. Different Escherichia coli B2-ST131 clades (B and C) producing extended-spectrum beta-lactamases (ESBL) colonizing residents of Portuguese nursing homes. *Epidemiology and infection* 2017;145(15):3303-3306.

33. **Che X, Chi F, Wang L, Jong TD, Wu CH et al.** Involvement of IbeA in meningitic Escherichia coli K1-induced polymorphonuclear leukocyte transmigration across brain endothelial cells. *Brain pathology (Zurich, Switzerland)* 2011;21(4):389-404.
34. **de Lorenzo V, Bindereif A, Paw BH, Neilands JB**. Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in Escherichia coli K-12. *Journal of bacteriology* 1986;165(2):570-578.
35. **Nipic D, Podlesek Z, Budic M, Crnigoj M, Zgur-Bertok D**. Escherichia coli uropathogenic-specific protein, Usp, is a bacteriocin-like genotoxin. *J Infect Dis* 2013;208(10):1545-1552.
36. **Nolan LK, Horne SM, Giddings CW, Foley SL, Johnson TJ et al.** Resistance to serum complement, iss, and virulence of avian Escherichia coli. *Veterinary research communications* 2003;27(2):101-110.
37. **Pramoonjago P, Kaneko M, Kinoshita T, Ohtsubo E, Takeda J et al.** Role of TraT protein, an anticomplementary protein produced in Escherichia coli by R100 factor, in serum resistance. *Journal of immunology (Baltimore, Md : 1950)* 1992;148(3):827-836.
38. **Russo TA, Wenderoth S, Carlino UB, Merrick JM, Lesse AJ**. Identification, genomic organization, and analysis of the group III capsular polysaccharide genes kpsD, kpsM, kpsT, and kpsE from an extraintestinal isolate of Escherichia coli (CP9, O4/K54/H5). *Journal of bacteriology* 1998;180(2):338-349.
39. **Sabri M, Leveille S, Dozois CM**. A SitABCD homologue from an avian pathogenic Escherichia coli strain mediates transport of iron and manganese and resistance to hydrogen peroxide. *Microbiology* 2006;152(Pt 3):745-758.
40. **Tu J, Xue T, Qi K, Shao Y, Huang B et al.** The irp2 and fyuA genes in High Pathogenicity Islands are involved in the pathogenesis of infections caused by avian pathogenic Escherichia coli (APEC). *Polish journal of veterinary sciences* 2016;19(1):21-29.
41. **Johnson TJ, Siek KE, Johnson SJ, Nolan LK**. DNA Sequence of a ColV Plasmid and Prevalence of Selected Plasmid-Encoded Virulence Genes among Avian Escherichia coli Strains. *Journal of bacteriology* 2005;188(2):745-758.
42. **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;84-85:11-19.
43. **Skyberg JA, Johnson TJ, Johnson JR, Clabots C, Logue CM et al.** Acquisition of avian pathogenic Escherichia coli plasmids by a commensal E. coli isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. *Infect Immun* 2006;74(11):6287-6292.
44. **Abraham S, Jordan D, Wong HS, Johnson JR, Toleman MA 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(4):273-277.
45. **Wyrsch E, Roy Chowdhury P, Abraham S, Santos J, Darling AE et al.** Comparative genomic analysis of a multiple antimicrobial resistant enterotoxigenic E. coli O157 lineage from Australian pigs. *BMC Genomics* 2015;16:165.
46. **Cain AK, Liu X, Djordjevic SP, Hall RM**. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant Salmonella enterica serovar Typhimurium from Australian animals. *Microbial drug resistance (Larchmont, NY)* 2010;16(3):197-202.

47. **Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP et al.** Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in Escherichia coli of human and animal origin. *PLoS One* 2010;5(9):e12754.
48. **Reid CJ, Roy Chowdhury P, Djordjevic SP**. Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. *Plasmid* 2015;80:127-137.
49. **Venturini C, Beatson SA, Djordjevic SP, Walker MJ**. Multiple antibiotic resistance gene recruitment onto the enterohemorrhagic Escherichia coli virulence plasmid. *FASEB J* 2010;24(4):1160-1166.
50. **McKinnon J, Roy Chowdhury P, Djordjevic SP**. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. *Int J Antimicrob Agents* 2018;52(3):430-435.
51. **Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM et al.** Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *Isme j* 2015;9(6):1269-1279.
52. **Harmer CJ, Hall RM**. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* 2016;1(2).
53. **Beaber JW, Hochhut B, Waldor MK**. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 2004;427(6969):72-74.
54. **Billman-Jacobe H, Liu Y, Haites R, Weaver T, Robinson L et al.** pSTM6-275, a Conjugative IncHI2 Plasmid of Salmonella enterica That Confers Antibiotic and Heavy-Metal Resistance under Changing Physiological Conditions. *Antimicrobial agents and chemotherapy* 2018;62(5).
55. **Curiao T, Canton R, Garcillan-Barcia MP, de la Cruz F, Baquero F et al.** Association of composite IS26-sul3 elements with highly transmissible IncI1 plasmids in extended-spectrum-beta-lactamase-producing Escherichia coli clones from humans. *Antimicrob Agents Chemother* 2011;55(5):2451-2457.
56. **Partridge SR, Tsafnat G, Coiera E, Iredell JR**. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol Rev* 2009;33(4):757-784.
57. **Fang L, Li X, Li L, Li S, Liao X et al.** Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from E. coli isolates of food-producing animals. *Sci Rep* 2016;6:25312.
58. **Whelan KF, Colleran E**. Restriction endonuclease mapping of the HI2 incompatibility group plasmid R478. *Journal of bacteriology* 1992;174(4):1197-1204.
59. **Barton MD**. Impact of antibiotic use in the swine industry. *Current opinion in microbiology* 2014;19:9-15.
60. **Chalmers G, Rozas K, Amachawadi R, Scott H, Norman K et al.** Distribution of the pco Gene Cluster and Associated Genetic Determinants among Swine Escherichia coli from a Controlled Feeding Trial. *Genes* 2018;9(10):504.
61. **Healy MG, Fenton O, Cormican M, Peyton DP, Ordsmith N et al.** Antimicrobial compounds (triclosan and triclocarban) in sewage sludges, and their presence in runoff following land application. *Ecotoxicology and environmental safety* 2017;142:448-453.
62. **Pritchard DL, Penney N, McLaughlin MJ, Rigby H, Schwarz K**. Land application of sewage sludge (biosolids) in Australia: risks to the environment and food crops. *Water science and technology : a journal of the International Association on Water Pollution Research* 2010;62(1):48-57.

63. **Jordan D, Chin JJ, Fahy VA, Barton MD, Smith MG et al.** Antimicrobial use in the Australian pig industry: results of a national survey. *Australian veterinary journal* 2009;87(6):222-229.
64. **Morales C, Lee MD, Hofacre C, Maurer JJ**. Detection of a novel virulence gene and a Salmonella virulence homologue among Escherichia coli isolated from broiler chickens. *Foodborne Pathog Dis* 2004;1(3):160-165.
65. **Sugimura K, Nishihara T**. Purification, characterization, and primary structure of Escherichia coli protease VII with specificity for paired basic residues: identity of protease VII and OmpT. *Journal of bacteriology* 1988;170(12):5625-5632.
66. **Jondreville C, Revy PS, Dourmad JY**. Dietary means to better control the environmental impact of copper and zinc by pigs from weaning to slaughter. *Livestock Production Science* 2003;84(2):147-156.

6.2.7 Figures

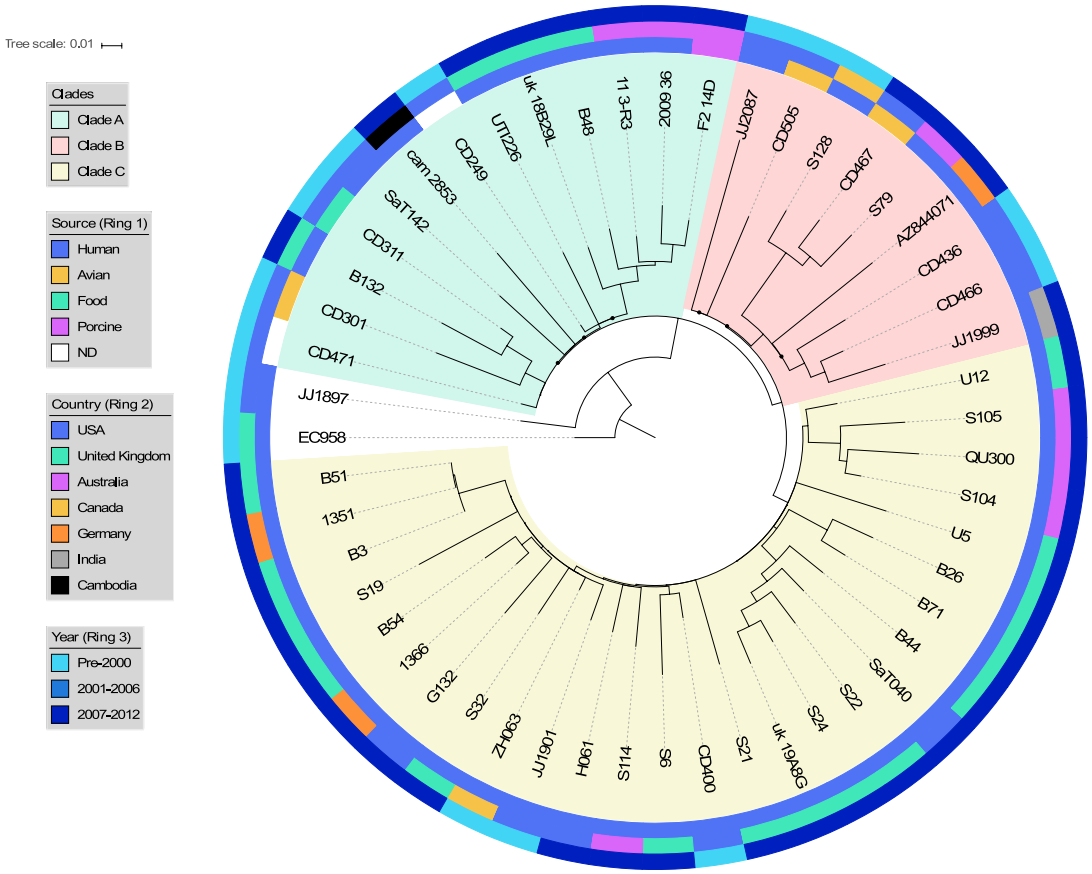


Figure 6.1

Maximum-likelihood tree of 50 ST131-*H*22 sequences derived from a core genome alignment of 2292 variable sites. Outer rings display metadata associated with each strain. Three major clades are shown. Black dots indicate nodes with confidence value <0.75.

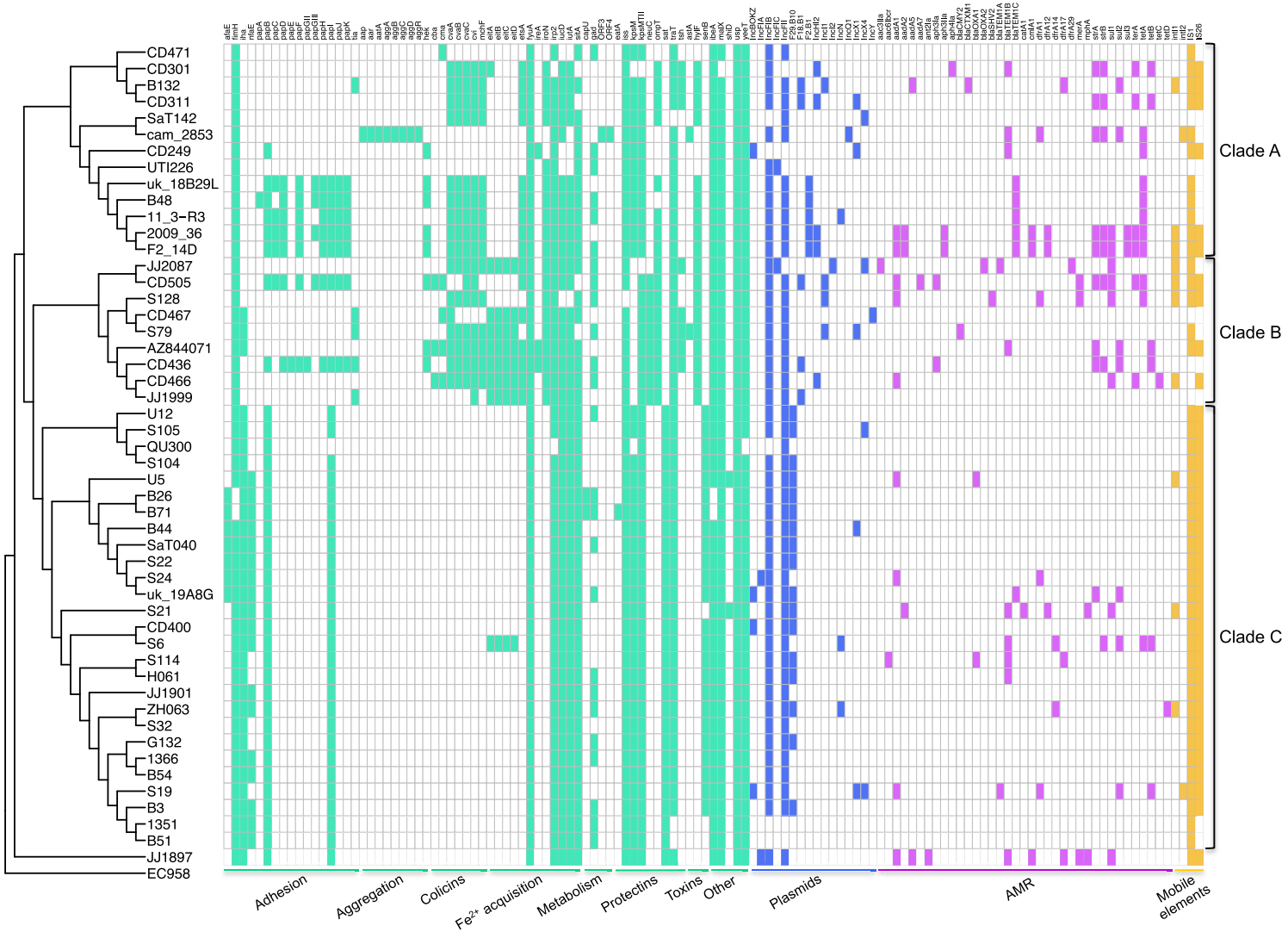


Figure 6.2

Maximum-likelihood tree of 50 ST131-*H*22 sequences mapped against VAG, plasmid-related, ARG and mobile element gene presence/absence. Colours indicate presence and white indicates absence. Green; VAGs, Blue; plasmid-related, Purple; ARGs, Yellow; mobile elements.

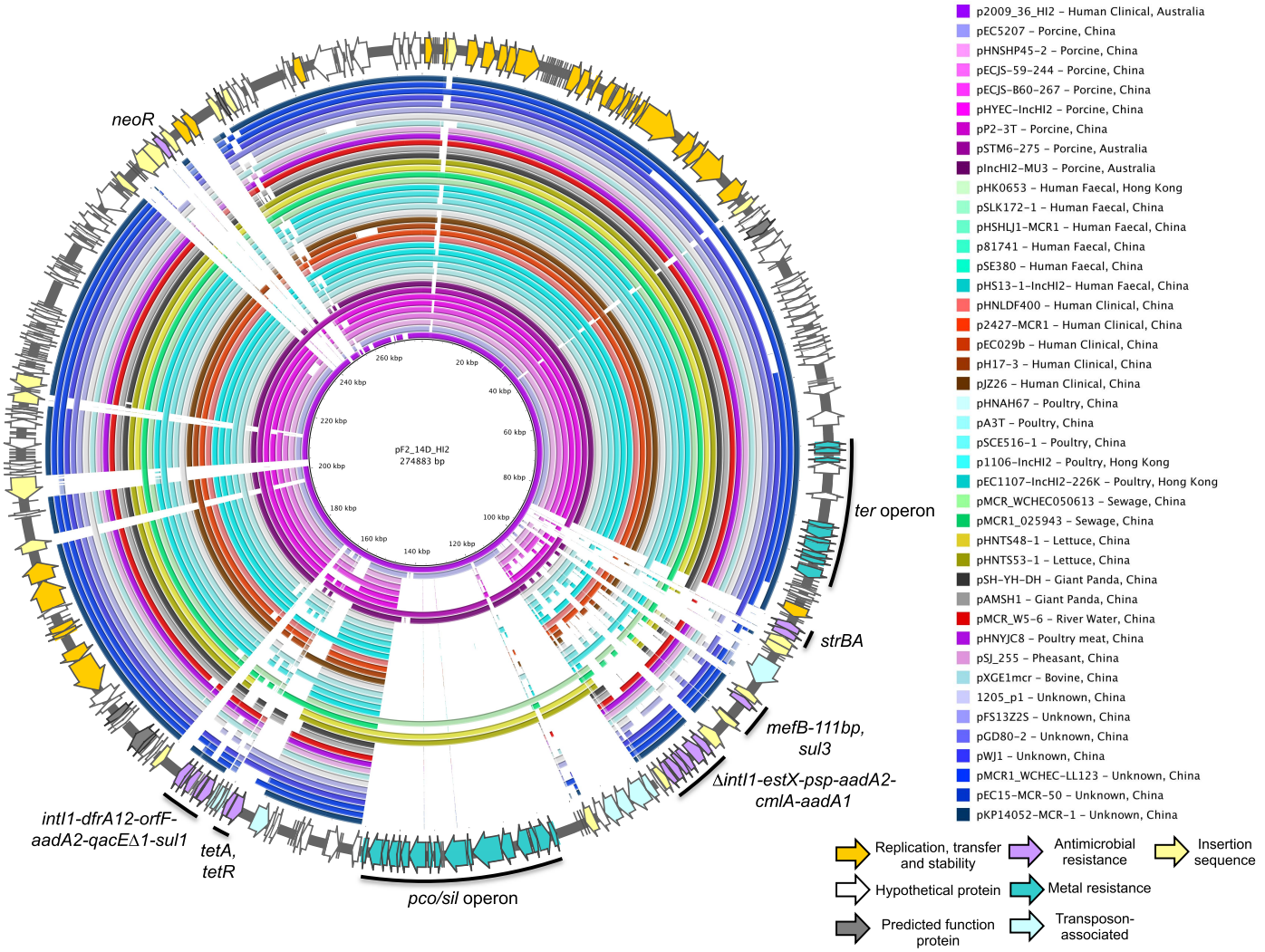


Figure 6.3

Schematic map of pF2_14D_HI2 (outer ring) and BRIG comparison to p2009_36_HI2 and 41 HI2-ST3 complete plasmids. Coloured arrows represent genes and their functions. Coloured rings indicate regions of sequence homology to pF2_14D_HI2 and white space indicates absence of homologous sequence.

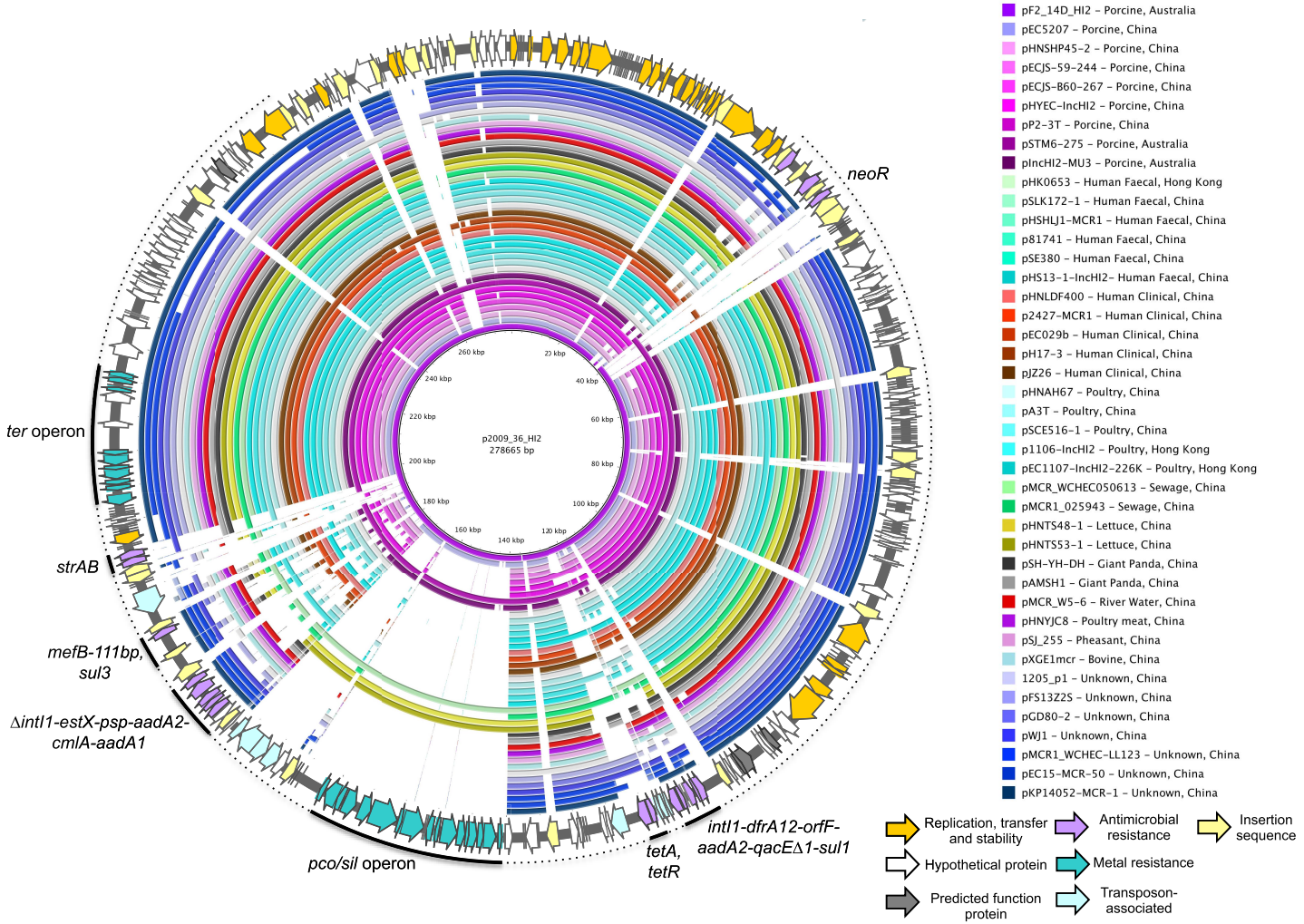


Figure 6.4

Schematic map of p2009_36_HI2 (outer ring) and BRIG comparison to pF2_14D and 41 HI2-ST3 plasmids downloaded from GenBank. Coloured arrows represent genes and their functions. Coloured rings indicate regions of sequence homology to p2009_36_HI2 and white space indicates absence of homologous sequence. The dotted line indicates the region that is inverted relative to pF2_14D_HI2.

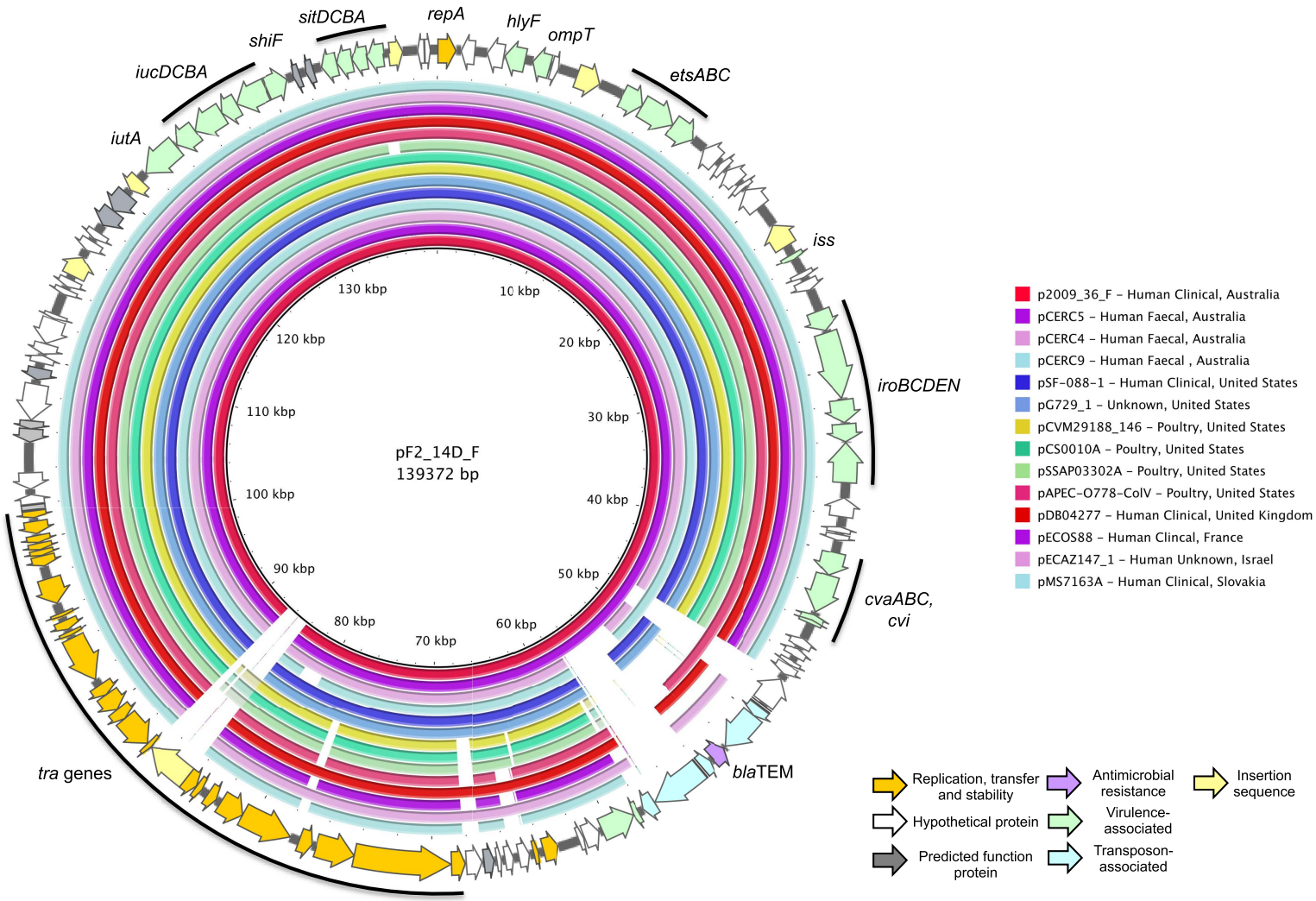


Figure 6.5

Schematic map of pF2_14D_F (outer ring) and BRIG comparison to p2009_36_F and 13 F2:B1 complete plasmids. Coloured arrows represent genes and their functions. Coloured rings indicate regions of sequence homology to pF2_14D_F and white space indicates absence of homologous sequence.

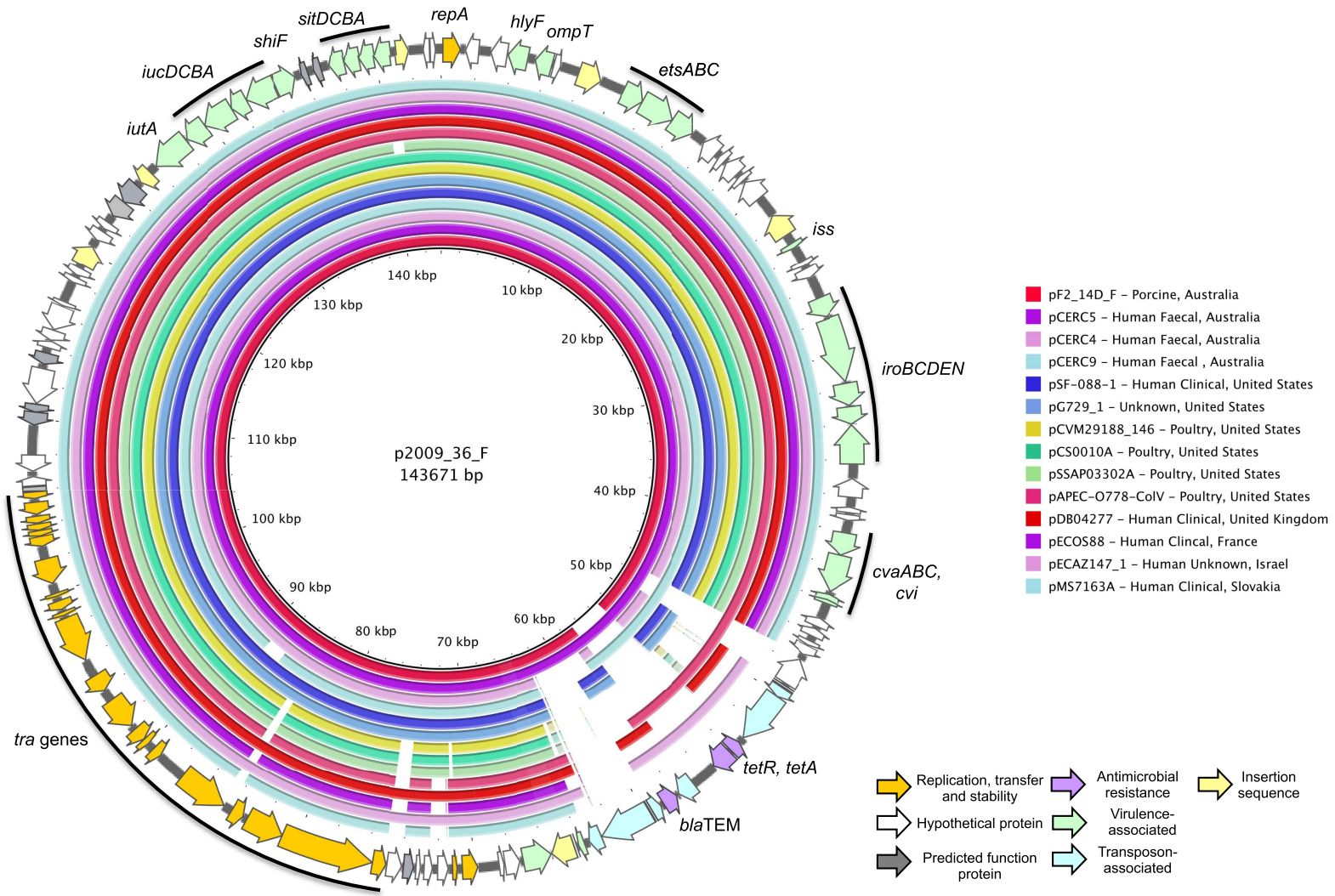


Figure 6.6

Schematic map of p2009_36_F (outer ring) and BRIG comparison to pF2_14D_F and 13 F2:B1 complete plasmids. Coloured arrows represent genes and their functions. Coloured rings indicate regions of sequence homology to p2009_36_F and white space indicates absence of homologous sequence.

Chapter 7: General discussion and future directions

As each results chapter contains discussion of the work therein, this final chapter will provide a general discussion based upon a synthesis of the aims, and discussion of some of the major implications and broader ideas arising from the thesis. Limitations will be addressed and future directions proposed.

Porcine faecal waste represents a significant environmental pollutant as it contains drug resistant microbes, heavy metals and antimicrobial residues. Unlike cattle and poultry, little is known about the genetic characteristics of commensal populations of porcine *E. coli,* which may pose a risk to both animal and human health. Prior to the commencement of this thesis whole genome sequence data of commensal MDR *E. coli* populations from swine in Australia was absent. Numerous PCR based studies mostly concerned pathogenic strains, identifying a limited set of VAGs and/or ARGs with very little consideration given to phylogenetic background, and plasmids[1-4]. In order to assess the risk of MDR porcine *E. coli* to human and animal health, all of these factors need to be taken into account together with the highest possible resolution. Establishment of databases of detailed genomic characteristics of *E. coli* from all areas of the biosphere will facilitate better tracking of important plasmids, the ARGs and VAGs they carry, and potentially pathogenic lineages. This in turn will allow identification of areas, such as food production systems, hospitals, municipal wastewater systems and natural waterways where strategies can be implemented to mitigate the evolution and spread of MDR *E. coli.* Currently, these efforts are in their infancy and individual populations first need to be characterised and understood while high-throughput whole genome sequencing and analysis pipelines are developed to cope with the enormous datasets that will be required to understand the problem on regional, national and global scales.

7.1 Aim 1

> *Determine the population structure, multi-locus sequence types, antimicrobial resistance genes, virulence-associated genes and plasmid types of a collection of multidrug resistant porcine commensal E. coli*

Aim 1 was primarily addressed in Chapter 4. This paper incorporated two collections of primarily commensal porcine MDR *E. coli* from two commercial swine production farms separated by approximately 250km. Sixty-eight commensal *E. coli* from 21 pigs on farm 2 (F2) were the primary focus of this thesis, whilst 35 farm 1 (F1) strains were the subject of another thesis and included 17 isolates from animals with diarrhoea. As stated in the manuscript, we considered the collection as a whole to be commensal due to the low carriage of ETEC virulence factors, which were present in both healthy and diseased pigs, as well as the fact that isolation of a strain from a disease animal does not confirm it as the causative pathogen[5].

The major findings of Chapter 4 included;

a) Identification of phylogroup A CC10, ST361 and ST542 as dominant among MDR commensal *E. coli* on two farms
b) Carriage of diverse integrons conferring multiple drug resistance
c) Carriage of ExPEC VAGs in commensal porcine *E. coli*

As these findings are discussed in Chapter 4 and many of the implications apply to Aim 2, this section will mostly address limitations within Chapter 4.

Firstly, the inclusion of only two farms was a clear limitation of the manuscript however, opportunistic sampling had to be accepted as a comprehensively designed genomic epidemiological study design incorporating many farms with a broad geographic spread was not practical at the time. Despite this limitation, a number of notable similarities were evident between the two farms such as ST overlap, carriage of identical integrons (D, G and H; Fig. 4.1 and Fig. 4.3) and plasmid replicons (IncF and IncHI2; Fig. 4.2). The similarity of resistance genotypes could be region specific and literature indicates AMR genotypes vary dramatically between farms due to differential selection pressures[4]. Therefore the conclusions about the presence and abundance of ARGs on these farms may not be applicable Australia-wide, they nonetheless highlight the need for genomic surveillance of AMR on a farm-to-farm basis.

In contrast to the AMR variability that is likely farm-specific, a number of sequence types observed on both farms are commonly reported as both commensals and pathogens in pigs worldwide[6-12]. Given this overlap, it seems that a broad range of mostly phylogroup A *E. coli* are well adapted to the pig gut and have been historically selected based on fitness prior to the additional pressure of antimicrobial use. It would be desirable to obtain sequence type data for multiple farms across Australia to determine how the national epidemiology of commensal *E. coli* compares to other countries and inform further studies that could elucidate the genetic basis of their success in pigs.

Whilst we were able to characterise a diversity of integrons, their context on plasmids and larger mobile elements was not determined due to the limitations of short read sequencing. Some strains carrying the same integron carried different plasmid replicons suggesting that the integrons or larger mobile scaffolds they reside on have transferred between plasmid types. Furthermore the presence of identical integrons and plasmids in different sequence types indicates that multiple mechanisms of HGT are at play in the gut of pigs. Whilst mercury resistance genes were common in the collection and known to be associated with integrons, we were unable to localise them on the same scaffold due to the limitations of short read sequencing. Further long read sequencing of a subset of representative plasmids, and subsequent read mapping of short read data back to these sequences is required to determine the context of integrons, the extent of integron transfer between plasmids, and plasmids between strains in the collection. Fully characterising these plasmids will allow them to be tracked in other populations and assess the contribution of pigs to the national and global burden of AMR.

The use of *intI1* as a proxy for MDR strains may be viewed as a limitation of the study. However, it should be noted that the broader collection from which the F2 strains originated was 98% PCR-positive for *intI1*. *intI1* selection is also preferable to using a single resistance gene or phenotype as a selection criteria because it allows examination of a greater variety of resistance genotypes. Furthermore, given the aim was to characterise MDR strains, as they pose the greatest risk to

human and animal health, we argue that *intI1* selection is appropriate and useful to this end[13]. One risk associated with this selection criteria is that truncated integrons, where *intI1* is interrupted by an IS element may be missed in PCR screening despite carrying cassette arrays and ARGs associated with other MGEs. IS-augmented integrons were present in the collection and are increasingly reported as drug resistance regions continually evolve[5, 14-16]. It is important that these potentially important structures are not overlooked. These insertion events are also useful as diagnostic markers for epidemiological tracking. Alternative strategies to select MDR strains with truncated integrons could include screening for IS elements or plasmid replicons that frequently carry ARGs.

7.2 Aim 2

> *Identify mobilised antimicrobial resistance genes, plasmids, clonal groups and strains and that may pose a risk to human and animal health*

This thesis identified two major characteristics of MDR porcine commensal *E. coli* that are cause for concern in both human and animal health. These included;

a) Diverse class 1 integrons augmented by IS*26,* associated with multiple plasmid types
b) Strains and lineages such as ST131 and CC10 that are likely human pathogens and carry multiple resistance determinants

The primary risk that porcine commensal *E. coli* pose to humans appears to be the contribution of successful plasmids carrying novel CRLs which comprise integrons, ARGs and IS elements that are primed for rapid evolution. Ten different class 1 integrons were identified in the collection, derived from five progenitor structures. Identical integrons were present in multiple STs and most of the variation among progenitors was due to IS*26*-mediated insertions and rearrangements. This is concerning for several reasons. Firstly, the distribution of successful resistance structures and plasmids among a broad range of *E. coli* lineages increases the chance of spread and success in multiple niches outside swine production where certain strains or sequence types might have an advantage. Secondly, the

modification of integrons by variable IS*26* insertions strongly supports the evolution of CRLs in swine production under antimicrobial pressure. This was particularly evident in integron (i) (Chapter 4), described on an IncHI2 plasmid in Chapter 6, which was abundant in the collection and has only been reported once on a highly related plasmid from a *Salmonella enterica* strain also originating from an Australian pig[17]. The subsequent selection of these novel genetic constructs and amplification of plasmids that carry them increases the chance of spillover and selection in other animal and human populations. This was apparent in the presence of a derivative of the IncHI2 plasmid in a human clinical ST131 strain described in Chapter 6. Finally, due to the utility of IS*26,* CRLs that carry multiple copies of IS*26* may become hotspots for the integration of new ARGs critical to human health. All that is required is the relevant selection pressure. As previously mentioned, more work needs to be done to fully characterise the diversity of plasmids and CRLs emerging from swine production so they can be tracked epidemiologically. IS*26* is paradoxically useful in this respect because the truncations it causes can be used as markers for tracking[18]. Whilst this thesis has characterised a number of resistance traits that may spread from one farm it is critical that this sort of surveillance is carried out on a national scale. The differential use of antimicrobials on different farms will certainly affect the plasmids and CRLs present[4]. A good example of this is the dissemination of $bla_{\text{CTX-M-1}}$ (conferring ceftiofur resistance) on an IncI1-ST3 plasmid on another Australian pig farm with a history of ceftiofur use[19].

Compounding the issue of evolution and amplification of CRLs and plasmids in swine production is their carriage in successful commensal and potentially pathogenic lineages. In Chapter 4 CC10 was identified as the predominant group of commensal *E. coli* in the strain collection and literature suggested CC10 and ST10 in particular are broad host range commensals as well as MDR opportunistic pathogens[5]. This lead to an exploration of how the porcine sequences compared to a global collection of CC10 strains. It was evident that CC10 strains colonizing pigs originate from multiple distinct lineages, supporting their broad fitness as commensals. Furthermore, given their VAG profiles and relatedness to strains from human blood and urine it is likely that some of the CC10 strains in our collection

are capable of extra-intestinal pathogenicity and could be considered ExPEC[20]. Whilst this would need to be confirmed in a model of infection, the genomic data provides strong support for this hypothesis. Further evidence of ST10 as a pathogen emerged in a recent report of colistin resistant pathogenic *E. coli* from pigs in Spain. The study observed molecular characteristics of three different pathotypes, ETEC, STEC and aEPEC, as well as hybrid pathotype STEC/ETEC among 29 *mcr1*-positive ST10 strains causing diarrhoea in pigs[12]. This raises a number of issues. Firstly, it is likely that most ST10 strains that cause disease in weaning piglets originate in their own gut and cause opportunistic infections when host immunity is compromised. Therefore ST10 poses a threat not only to human health but animal health as well. Secondly, the combination of colistin resistance and MDR phenotypes within these ST10 strains are a serious cause for concern should these strains or plasmids transfer to humans where colistin is a last-line treatment option for gram-negative infections. Though colistin is not used in animal production in Australia, this study serves as an example of the serious problems that can occur when successful, diverse and potentially pathogenic *E. coli* in the gut of animals are exposed to last-line antimicrobials. The potential role of swine production in driving evolution and dissemination of AMR was also highlighted by the distinct ARG and plasmid carriage of our porcine CC10 strains relative to the rest of the global collection.

The identification of an ST131 strain carrying a ColV virulence plasmid and a locally acquired IncHI2 resistance plasmid was a surprising example of a pathogen present in the commensal microbiome of a pig. Its striking similarity to a human clinical strain all but confirms its identity as a genuine pathogen and its carriage of resistance that seemingly evolved in swine production is concerning. Despite the global focus on ESBL carriage in ST131-*H*30*Rx* from humans, increasing evidence suggests that the *H*22 precursor is well adapted to the gut of animals and readily acquires drug resistance where pressure is present[12, 21]. It will be interesting to see if *H*22 strains with antimicrobial resistance characteristics indicative of animal production continue to emerge in humans.

In addition to the ST131 strain, two highly virulent ST117 strains were identified in Chapter 4. ST117 is a common pathogen of poultry and known ExPEC in humans [22-24]. These strains did not carry an IncHI2 replicon, however their resistance genes were similar to many other strains in the collection indicating they have also acquired resistance traits within the farm. We intend to perform long read sequencing of these strains in order to fully characterise the plasmids they carry.

A clear limitation of both chapters 5 and 6 was the use of publicly available sequences for comparison and our inability to control for any epidemiological characteristics. This may be the reason there was no clear congruence between the genomic characteristics of strains and their epidemiological metadata in either CC10 or ST131-*H*22. There was a contrast in the diversity of CC10 compared to ST131-*H*22, which is clearly a more clonal population. Future efforts to map and model the dissemination and transfer of clonal groups and AMR between humans, animals and the environment on any scale need to take into account this variation. Highly clonal groups like ST131 will potentially be easier to track and understand, coupled with their relevance to human and animal health, these groups need to be understood first and the lessons learnt carried over to analysis of more diverse lineages such as CC10. The expansion of global databases will assist in this regard, however new methods of analysis and mathematical models will need to be developed in order to truly understand the situation. It would be beneficial for public databases to establish a minimum standard of metadata required when submitting genome sequences to increase the power of genomic epidemiological studies.

7.3 Aim 3

Use long read sequencing to fully characterise strains and plasmids identified in Aim 2 and compare them to relevant human origin strains

The similarity between porcine ST131 strain F2_14D and human clinical strain 2009_36 observed from Illumina data led to us performing long read sequencing to fully characterise and compare the plasmids they carried. Both strains carried remarkably similar ColV and IncHI2 plasmids (pF2_14D_F, p2009_36_F,

pF2_14D_HI2 and p2009_36_HI2), supporting their shared evolutionary history. Both of these plasmids are relevant to human health and their presence in pig production is concerning. Due to the fitness and virulence-associated characteristics of ColV plasmids, they are worthy of further investigation in the gut of animals where they may be playing a role in the emergence of potentially zoonotic pathogens in conjunction with antimicrobial resistance[25]. A number of other strains in the original collection exhibited genotypes suggestive of ColV carriage and further analysis is required to determine if they carry plasmids related to pF2_14D_F and if they also carry ARGs. We hypothesise that this plasmid or closely related derivatives have transferred between different STs on the farm.

The carriage of an apparently porcine origin IncHI2 plasmid in a human ExPEC strain is a clear example of the risk swine production could pose to human health. Furthermore, the evolution and rearrangements driven by IS*26* in the human origin plasmid highlight how the evolution of IS*26*-associated resistance regions in animal production can continue to evolve in humans and hospitals. Data in Chapter 4 indicated that similar plasmids are present in many of our other porcine strains and mapping of short reads to pF2_14D_HI2 is needed to determine how similar they are. It was notable that related plasmids carrying integrons and ARGs are abundant in animals in China, indicating a regional distribution of this particular lineage[26, 27]. Potential modes of microbial and AMR transfer between countries and their food production systems should be investigated in future.

As was the case in examining global CC10 and ST131 strains, a limitation of this analysis was the use of publicly available complete plasmid sequences, which are less abundant than Illumina sequence data. We were able to compare a number of IncHI2-ST3 plasmids, which indicated an Asia-Pacific distribution. However, it is still possible these plasmids are present in other parts of the world and have simply not been sequenced. Ever expanding databases and large-scale analyses will shed light on the global distribution of resistance plasmids in the future.

7.4 Future directions

In assessing the risks pig production poses to human health we have only sampled a very small part of the picture. This thesis provides a basis for further analysis that will yield stronger conclusions, though it is remarkable what has emerged from such a small-scale analysis.

There is scope for further work on this collection as well as the field of genomic epidemiology more broadly. Firstly, long read sequencing of more strains within CC10 as well as the ST117 strains should be performed to fully characterise their resistance plasmids. It will be interesting to see if they carry virulence plasmids, which contribute to their risk to human health. As a major focus of the thesis was AMR, it was somewhat surprising that mobile virulence plasmids were present. The potential evolution and selection of virulence plasmids that confer pathogenicity in otherwise non-virulent, commensal strains in animal production is problematic and should be closely monitored together with AMR plasmids. Co-carriage of virulence and resistance plasmids rich in IS elements increases the chance that hybrid virulence-resistance plasmids will evolve. Such a situation could lead to pathogenicity being selected as a by-product of antimicrobial use. Numerous examples of hybrid virulence and resistance plasmids in humans and animals are reported and important in terms of genomic surveillance[25, 28, 29].

Secondly, the scale of sampling in pig production in Australia needs to be expanded. A 2016 study isolated 5003 *E. coli* strains from swine on 72 farms in Australia[4]. Whole genome sequencing of a collection of that scale would dramatically increase understanding of the risk porcine commensal *E. coli* pose to human and animal health in Australia. This will require strong collaboration between government, scientific institutes, veterinarians, farmers and animal production corporate bodies with the common goal of increasing biosecurity.

In addition to expanding sampling in swine production, genomic epidemiology studies need to sample temporally and spatially related strains, unbiased by selection of specific antimicrobial resistance phenotypes from a wide variety of settings in order to understand relationships between reservoirs. It would be beneficial to examine a rural area where a variety of agricultural production settings are in close proximity to humans. A study might involve sampling:

1. Hospitals
    a. Patients; infectious and commensal sites
    b. Medical staff
    c. Physical surfaces such as sinks, drains and beds
    d. Wastewater and effluent
2. Community
    a. Food from supermarkets and restaurants
    b. Faecal flora of healthy individuals
    c. Public spaces
3. Agriculture
    a. Animal faeces
    b. Farmers
    c. Wastewater
    d. Irrigation water
    e. Fresh produce where animal manure is used
    f. Soil
4. Water
    a. Natural waterways, particularly those in proximity to agriculture
    b. Stormwater
    c. Sewage
    d. Treated wastewater

Such a comprehensive genomic dataset of *E. coli* would be of huge benefit, not just for the area sampled but global genomic epidemiology on the whole. Dorado-Garcia et al performed a similar study, however it was significantly biased by selection of ESBL/AmpC-producing *E. coli.* Selection for last-line antimicrobial resistance phenotypes ignores the broader factors such as successful plasmid

types, antimicrobial resistance transposons and other MGEs such as IS elements that facilitate their emergence. It is impossible to understand the relative contribution of animals to AMR in humans with such an approach. The nature of the problem of AMR is still far from being fully characterised or understood. There is a desperate need to understand the dynamics of strain and AMR movement between all of the above settings so that well-designed interventions can be implemented. A phenotypically unbiased study such as the one described might inform the development of models and analysis methods that can be applied to other towns, cities and on national and global scales.

The need for models that characterise the dynamics of transfer goes hand in hand with the need for development of methods of analysis that can deal with the sheer scale of data that will be generated in the future. Increased computing power, new mathematical models and standardized automated analysis pipelines are particularly critical in this regard.

7.5 References

1. **Abraham S, Jordan D, Wong HS, Johnson JR, Toleman MA 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(4):273-277.
2. **Chapman TA, Wu XY, Barchia I, Bettelheim KA, Driesen S et al.** Comparison of virulence gene profiles of Escherichia coli strains isolated from healthy and diarrheic swine. *Appl Environ Microbiol* 2006;72(7):4782-4795.
3. **Smith MG, Jordan D, Chapman TA, Chin JJ, Barton MD et al.** Antimicrobial resistance and virulence gene profiles in multi-drug resistant enterotoxigenic Escherichia coli isolated from pigs with post-weaning diarrhoea. *Veterinary microbiology*, Research Support, Non-U.S. Gov't 2010;145(3-4):299-307.
4. **Smith MG, Jordan D, Gibson JS, Cobbold RN, Chapman TA et al.** Phenotypic and genotypic profiling of antimicrobial resistance in enteric Escherichia coli communities isolated from finisher pigs in Australia. *Aust Vet J* 2016;94(10):371-376.
5. **Reid CJ, Wyrsch ER, Roy Chowdhury P, Zingali T, Liu M et al.** Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. *Microb Genom* 2017;3(12).
6. **Ahmed S, Olsen JE, Herrero-Fresno A**. The genetic diversity of commensal Escherichia coli strains isolated from non-antimicrobial treated pigs varies according to age group. *PLoS One* 2017;12(5):e0178623.
7. **Bednorz C, Oelgeschlager K, Kinnemann B, Hartmann S, Neumann K et al.** The broader context of antibiotic resistance: zinc feed supplementation of piglets increases the proportion of multi-resistant Escherichia coli in vivo. *Int J Med Microbiol* 2013;303(6-7):396-403.
8. **Cortes P, Blanc V, Mora A, Dahbi G, Blanco JE et al.** Isolation and characterization of potentially pathogenic antimicrobial-resistant Escherichia coli strains from chicken and pig farms in Spain. *Applied and environmental microbiology* 2010;76(9):2799-2805.
9. **Garcia V, Garcia-Menino I, Mora A, Flament-Simon SC, Diaz-Jimenez D et al.** Co-occurrence of mcr-1, mcr-4 and mcr-5 genes in multidrug-resistant ST10 Enterotoxigenic and Shiga toxin-producing Escherichia coli in Spain (2006-2017). *Int J Antimicrob Agents* 2018.
10. **Herrero-Fresno A, Larsen I, Olsen JE**. Genetic relatedness of commensal Escherichia coli from nursery pigs in intensive pig production in Denmark and molecular characterization of genetically different strains. *J Appl Microbiol* 2015;119(2):342-353.
11. **Wang J, Gibbons JF, McGrath K, Bai L, Li F et al.** Molecular characterization of blaESBL-producing Escherichia coli cultured from pig farms in Ireland. *J Antimicrob Chemother* 2016.
12. **Garcia-Menino I, Garcia V, Mora A, Diaz-Jimenez D, Flament-Simon SC et al.** Swine Enteric Colibacillosis in Spain: Pathogenic Potential of mcr-1 ST10 and ST131 E. coli Isolates. *Front Microbiol* 2018;9:2659.
13. **Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM et al.** Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *Isme j* 2015;9(6):1269-1279.

14. **Chen X, He L, Li Y, Zeng Z, Deng Y et al.** Complete sequence of a F2:A-:B- plasmid pHN3A11 carrying rmtB and qepA, and its dissemination in China. *Vet Microbiol* 2014;174(1-2):267-271.
15. **Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP et al.** Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in Escherichia coli of human and animal origin. *PLoS One* 2010;5(9):e12754.
16. **Wang W, Baloch Z, Peng Z, Hu Y, Xu J et al.** Genomic characterization of a large plasmid containing a bla NDM-1 gene carried on Salmonella enterica serovar Indiana C629 isolate from China. *BMC infectious diseases* 2017;17(1):479.
17. **Billman-Jacobe H, Liu Y, Haites R, Weaver T, Robinson L et al.** pSTM6-275, a Conjugative IncHI2 Plasmid of Salmonella enterica That Confers Antibiotic and Heavy-Metal Resistance under Changing Physiological Conditions. *Antimicrobial agents and chemotherapy* 2018;62(5).
18. **Wyrsch ER, Roy Chowdhury P, Chapman TA, Charles IG, Hammond JM et al.** Genomic Microbial Epidemiology Is Needed to Comprehend the Global Problem of Antibiotic Resistance and to Improve Pathogen Diagnosis. *Front Microbiol* 2016;7:843.
19. **Abraham S, Kirkwood RN, Laird T, Saputra S, Mitchell T et al.** Dissemination and persistence of extended-spectrum cephalosporin-resistance encoding IncI1-blaCTXM-1 plasmid among Escherichia coli in pigs. *The ISME Journal* 2018;12(10):2352-2362.
20. **Reid CJ, DeMaere MZ, Djordjevic SP**. Australian porcine clonal complex 10 (CC10) Escherichia coli belong to multiple sublineages of a highly diverse global CC10 phylogeny. *Microb Genom* 2018.
21. **Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K et al.** Escherichia coli ST131-H22 as a Foodborne Uropathogen. *mBio* 2018;9(4).
22. **Mora A, Lopez C, Herrera A, Viso S, Mamani R et al.** Emerging avian pathogenic Escherichia coli strains belonging to clonal groups O111:H4-D-ST2085 and O111:H4-D-ST117 with high virulence-gene content and zoonotic potential. *Vet Microbiol* 2012;156(3-4):347-352.
23. **Manges AR, Harel J, Masson L, Edens TJ, Portt A et al.** Multilocus sequence typing and virulence gene profiles associated with Escherichia coli from human and animal sources. *Foodborne Pathog Dis* 2015;12(4):302-310.
24. **Ronco T, Stegger M, Olsen RH, Sekse C, Nordstoga AB et al.** Spread of avian pathogenic Escherichia coli ST117 O78:H4 in Nordic broiler production. *BMC Genomics* 2017;18(1):13.
25. **Moran RA, Hall RM**. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. *Microbial drug resistance (Larchmont, NY)* 2017.
26. **Fang L, Li X, Li L, Li S, Liao X et al.** Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from E. coli isolates of food-producing animals. *Sci Rep* 2016;6:25312.
27. **Zhao H, Chen W, Xu X, Zhou X, Shi C**. Transmissible ST3-IncHI2 Plasmids Are Predominant Carriers of Diverse Complex IS26-Class 1 Integron Arrangements in Multidrug-Resistant Salmonella. *Front Microbiol* 2018;9:2492.
28. **Venturini C, Beatson SA, Djordjevic SP, Walker MJ**. Multiple antibiotic resistance gene recruitment onto the enterohemorrhagic Escherichia coli virulence plasmid. *FASEB J* 2010;24(4):1160-1166.

29. **Venturini C, Hassan KA, Roy Chowdhury P, Paulsen IT, Walker MJ et al.** Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS26-related molecular signature isolated from different Escherichia coli pathotypes from different hosts. *PLoS One* 2013;8(11):e78862.

Appendix

File S6.1 SDS/Phenol Chloroform Extraction Protocol

Before starting:

1. Set heat blocks to 37 and 56 degrees C.
2. Check you have the following reagents/solutions
    a. TE
    b. 20% SDS
    c. Proteinase K
    d. RNase A
    e. Ethanol 70% and 100%
    f. Phenol/chloroform/isoamyl alcohol
    g. Chloroform/isoamyl alcohol
    h. 3M sodium acetate pH5.2

1. Prepare an overnight culture in rich media.

2. Sub-culture 100uL of this overnight culture in 10mL fresh media and incubate for 3-3.5 hours.

3. Spin mid-log cells for 15 minutes at 4000 rpm and re-suspend in 500uL of Lysis buffer:

482uL	TE (pH8.0)
15uL	20% SDS
3uL	Proteinase K (20mg/mL)

4. Incubate this tube at 56°C with mixing via tube inversion every 15 minutes until completely lysed.

5. Add 1/10 volume (50ul) 3M sodium acetate pH5.2 and 2.5x volume (1.25mL) 100% ethanol and mix gently.

6. With a pipette tip, transfer the precipitated proteins and nucleic acids into a tube containing 500uL 70% ethanol.

7. Spin tubes at 13000 rpm for 10 minutes, pour off supernatant. Resuspend pellet in 800uL water (ideally pre-heated to 65°C).

8. Add 10uL of RNase A, incubate at 37°C for 30 minutes.

9. Bring the tube back to room temperature after incubation and move on with Phenol-Chloroform extraction.
10. Pre-mix phenol/chloroform/isoamyl alcohol solution (400uL of phenol; 400uL chloroform/isoamyl alcohol) and add (total 800uL).
11. Invert 6-8 times then spin at 13000 rpm for 5 minutes.
12. Remove as much liquid as possible from the top aqueous phase and place into a new tube. Avoid picking up any of the phenol/chloroform/isoamyl alcohol phase.
13. Add equal volumes of the chloroform/isoamyl alcohol solution to extracted aqueous phase.
14. Invert 6-8 times then spin at 13000 rpm for 5 minutes.
17. Remove as much of the top aqueous phase as possible and place into a new tube. Avoid picking up any of the chloroform/isoamyl alcohol phase.
18. Repeat steps 5 and 6 (ethanol precipitation, adjust volume of sodium acetate and ethanol based on how much aqueous phase you get).
19. Spin for 20 minutes at 13000 rpm (optionally at 4°C).
20. Decant supernatant carefully without disturbing the pellet.
21. Wash by adding 500uL of 70% ethanol and invert 6-8 times.
22. Spin for 15 minutes at 13000 rpm (optionally at 4°C).
23. Decant supernatant carefully without disturbing the pellet.
24. Repeat the wash steps (steps 21-23)
25. Quick spin on table top centrifuge to draw residual ethanol to bottom.
26. Remove residual ethanol with a p20 pipette, be careful not to disturb the pellet.
27. Air dry pellet and resuspend with appropriate volume (100uL) of pre-heated 65°C water.