Whole genome sequence analysis of Avian Pathogenic *E. coli* to assess their potential as reservoirs of antimicrobial resistance and as potential zoonotic pathogens.

Max Laurence Cummins

Bachelor of Biotechnology (Hons)

University of Technology Sydney

Faculty of Science

School of Life Sciences

The i3 Institute

September 2020

Certificate of Authorship

I, Max Laurence Cummins declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science, School of Life Sciences at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced 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 the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: **29/09/2020**

Acknowledgements

Firstly, I would like to thank my supervisor Steve, for whom I have a tremendous amount of respect and gratitude. Your contagious enthusiasm has been an endless source of motivation over the years and I greatly appreciate your support and mentorship. I will fondly remember the conference trips I shared with you and other members of our lab, particularly our trip to Italy. Thanks to Tiz for helping make that trip so special for all of us, as well as for her comradery and general role as life coach during her and my time studying together. Mohammad, you played a critical role in guiding me through the development of much of this thesis – thanks a lot!

Thank you to our collaborators in Australia and abroad, particularly Marc Marenda and Glenn Browning for their help with the collection which I studied for most of my thesis. Thanks also to Ausgem, UTS and the i3 Institute for funding and hosting my research, and the Australian Government Research Training Program.

Also, a big thank you to Piklu, who helped greatly in showing me the ropes in the lab, and also Ethan, who helped me learn some of the computational aspects of my research. Thanks to Veronica, I've found your relentless work ethic to be something to aspire to. Thanks to Dmitriy, I have been grateful to have your companionship and have enjoyed our nerdy chats about analytical pipelines and videogames!

I would also like to thank my colleague and dear friend Cameron. I have learnt a lot from you, as a person and an academic, and I look forward to us both being old and presumably bald (or at least balding) bioinformatician types when our time comes.

Thanks also to Carla for helping me through some difficult times during my studies.

Lastly, I would like to thank my family. Thank you to my parents for letting me rehearse my presentations to you, listening to my rants at the dinner table, for your financial and emotional support and for always reminding me how proud you are of me. Thanks to my grandparents for helping me through tough times and for their kindness and tremendous generosity over the years. Thanks to my brother Elliot for always taking an interest in my work and for always having been there for me when I need him. I love you all, thank you for helping me get here.

Statement

This thesis is by publication. Chapters 4, 5 and 6 constitute the results chapters and each has been published in a peer reviewed journal. Please note that numbers of figure, table and sections within these chapters have been changed to maintain consistency and flow with the thesis document as a whole. Figures from each results chapter have been included as additional files and readers may prefer to view these rather than those present in this document.

List of publications

Paper 1 - Chapter 4

Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene

Max L. Cummins[1], Cameron J. Reid[1], Piklu Roy Chowdhury[1], Rhys N. Bushell[2], Nicolas Esbert[2], Kelly A. Tivendale[2], Amir H. Noormohammadi[2], Shaiful Islam[2], Marc S. Marenda[2], Glenn F. Browning[2], Philip F. Markham[2] and Steven P. Djordjevic[1*]

Published in Microbial Genomics 23/01/2019: https://doi.org/10.1099/mgen.0.000250.

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

[2] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, and Werribee, Victoria 3030, Australia.

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Paper 2 - Chapter 5

Salmonella Genomic Island 1B Variant Found in a Sequence Type 117 Avian Pathogenic *Escherichia coli* Isolate

Max Laurence Cummins[a], Piklu Roy Chowdhury[a], Marc Serge Marenda[c], Glenn Francis Browning[b], Steven Philip Djordjevic[a*]

Published in mSphere 22/05/2019: https://doi.org/10.1128/mSphere.00169-19.

[a] The ithree institute, University of Technology Sydney, Ultimo, New South Wales, Australia

[b] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[c] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Werribee, Victoria, Australia

* Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Paper 3 - Chapter 6

Salmonella Genomic Island 1 is Broadly Disseminated within Gammaproteobacteriaceae

Max L. Cummins[1,2], Mohammad Hamidian[1] and Steven P. Djordjevic[1,2*]

Published in Microorganisms 20/01/2020:
https://doi.org/10.3390/microorganisms8020161

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

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

* Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Table of Contents

Abbreviations

3`-CS	3`-conserved segment
5`-CS	5`-conserved segment
AMR	Antimicrobial resistance
AMS	Antimicrobial susceptible
APEC	Avian pathogenic *E. coli*
cgMLST	Core-genome multi-locus sequence type
CIA	Critically important antimicrobial
ESBL	Extended-spectrum beta-lactamase
ExPEC	Extraintestinal pathogenic *E. coli*
FUTI	Foodborne-urinary tract infection
HC	Hierarchical cluster
HierCC	Hierarchical Clustering of cgMLST
IME	Integrative mobilisable element
IPEC	Intestinal pathogenic *E. coli*
IS	Insertion sequence
KGI	*Klebsiella* genomic island
LB	Lysogeny broth
LGT	Lateral gene transfer
MDR	Multidrug resistant
MGE	Mobile genetic element
MLST	Multi-locus sequence type
NMEC	Neonatal meningitis-associated *E. coli*
ONT	Oxford Nanopore Technologies
ORF	Open reading frame

PCR	Polymerase chain reaction
PGI1/2	*Proteus* genomic island
SGI1-RE	*Salmonella* genomic island 1-related element
SGI1/2	*Salmonella* genomic island 1/2
SNP	Single nucleotide polymorphism
SRA	Sequence read archive
ST	Sequence type
STEC	Shiga-toxigenic *E. coli*
Tn	Transposon
UPEC	Uropathogenic *E. coli*
UTI	Urinary-tract infection
VAG	Virulence-associated gene
VGI	*Vibrio* genomic island
WGS	Whole genome sequencing
XDR	Extensively drug resistant

Abstract

Antimicrobial resistance (AMR) is a complex, global health challenge requiring a One Health framework, acknowledging the interconnectedness of human, animal and environmental microbiospheres, in order to understand its reservoirs, transfer mechanisms and evolution. *Escherichia coli* is a model 'One Health organism' as it is a gastrointestinal commensal of diverse vertebrate species, a common contaminant of foodstuffs and natural environments, increasingly multidrug resistant (MDR) and a major human pathogen.

Within humans, extraintestinal pathogenic *E. coli* (ExPEC) cause severe disease, including urinary tract infections (UTIs) and blood sepsis. Similarly, avian pathogenic *E. coli* (APEC), a subset of ExPEC, cause significant economic losses in the poultry sector. Despite known phylogenetic and genotypic similarities between the two, it is still not possible to define them with genomic data alone and their transit between poultry and humans is poorly characterised. Genomic data on Australian APEC in particular is lacking, hampering efforts to assess and mitigate their threat to human and animal health.

Accordingly, we characterised the genomes of 95 Australian MDR APEC sourced from four Australian states. ST117 and clonal-complex 350 were abundant, however pandemic human ExPEC lineages also featured, including ST95, ST73 and ST131. Almost all strains carried large ColV-like virulence plasmids, however no strains carried genes conferring resistance to critically important antimicrobials (CIAs), such as carbapenems, extended-spectrum β–lactams, fluoroquinolones and colistin, except one exhibiting a fluoroquinolone-resistance associated mutation. This attests to the value of strict antimicrobial use regulations in food animals in Australia.

One strain carried a novel variant of *Salmonella* genomic island 1 (SGI1) named SGI1-B-Ec1. SGI1 is an integrative mobilisable element that confers resistance to five different classes of antimicrobials and may confer pathogenicity traits. It was originally characterised in outbreak strains of *Salmonella enterica* serovar Typhimurium, however SGI1-related elements (SGI1-REs) have been reported in diverse genera, some carrying CIA-resistance genes. The first identification of an SGI1-RE in *E. coli* is alarming as it indicates other more clinically significant SGI1-REs may exist within Australian food-

animals. Subsequently, high-throughput analysis of 455,632 bacterial genomes revealed previously undescribed SGI1-REs in *E. coli, Klebsiella pneumoniae, Vibrio cholerae* and *Cronobacter sazakii*.

This data provides valuable insight into the current AMR status of Australian MDR APEC, as well as the spread and evolution of SGI1 and SGI1-REs within Gammaproteobacteriaceae. Large-scale, One Health-oriented genomic epidemiological studies are urgently required into APEC and bacterial populations more broadly to identify and mitigate their threat to human and animal health.

Chapter 1: Thesis Overview

1.1 Overview

This research project involved analysis of the genomic sequence data of 95 isolates of multi-drug resistant isolates of avian pathogenic *E. coli* (APEC) originating from diseased poultry around Australia. During this analysis, one isolate was found to carry a variant of *Salmonella* genomic island 1 (SGI1), a genetic element important in the evolution and spread of AMR in Enterobacteriaceae. This was the first report of SGI1 within *E. coli*, a finding that prompted the analysis of large online databases for further evidence of SGI1 and SGI1-related elements within other genera previously not reported to carry them. All three chapters of this thesis have now been published.

1.2 Aims

- Aim 1: to determine the phylogenetic characteristics, virulence- and AMR-associated gene content of a geographically-diverse collection of Australian MDR APEC

  Hypotheses:

  - Australian MDR APEC do not frequently carry genes conferring resistance to critically important antimicrobials
  - These isolates are rich in plasmid-encoded, virulence-associated genes (VAGs), particularly those involved in iron acquisition
  - A subset of these APEC exhibit phylogenetic and genotypic similarity with human ExPEC

• Aim 2: to characterise mobile genetic elements (MGEs) associated with Australian MDR APEC

  Hypotheses:

  - Australian MDR APEC frequently carry ColV-like virulence plasmids
  - Australian MDR APEC exhibit various genetic contexts for AMR determinants and elements important in AMR evolution and dissemination

• Aim 3: to assess the spread of SGI1 and SGI1-REs throughout publicly available genomic datasets

Hypothesis:

- SGI1 and SGI1-REs have a more extensive host range than has previously been reported in the literature

1.3 Summary and contribution to the literature

This thesis provides whole genome sequence (WGS) data on an Australian population of MDR APEC, a cause of major disease burden within the poultry industry and potential contributor to human infections. Analysis revealed the phylogenetic composition of APEC throughout various Australian states, described the antimicrobial resistance genotypes common in these organisms, and assessed their VAG carriage. Additionally, this thesis reports the first finding of an SGI1 variant in a strain of *E. coli*, and also provides novel data on the occurrence of SGI1 and SGI1-REs in other Gammaproteobacteriaceae.

The primary findings include:

- ST117 is likely the most dominant sequence type of APEC throughout Australian poultry operations
- Australian MDR APEC are not a major reservoir for genes conferring resistance to critically important antimicrobials (CIAs)
- *E. coli, Vibrio cholerae, Klebsiella pneumoniae* and *Cronobacter sazakii* have acquired SGI1-REs

Chapter 2: Literature Review

2.1 Antimicrobial resistance

Antimicrobial resistance (AMR) is a global health challenge considered by the World Health Organisation (WHO) as both urgent and complex (1). AMR refers to the capacity of a microorganism to resist the effects of antimicrobial therapeutics. The primary impacts of AMR on human populations are increased morbidity, mortality, length of hospitalisation and healthcare expenditure. For example, a study in the United States on a cohort of 1,400 hospital patients found that AMR infections, in contrast with antimicrobial susceptible (AMS) infections, were attributed to a 6.5% increase in mortality, a 1-2 week increased length of hospital stay and an increased treatment cost of $18,000 USD (2).

Antimicrobial resistant pathogens are increasing in prevalence, with such infections in the United States having doubled between 2002 and 2014 from 5.2% to 11% (3). Multiple drug resistant (MDR) strains, especially those resistant to the critically important antimicrobials (CIAs) 3rd, 4th and 5th generation cephalosporins, glycopeptides, macrolides, polymyxins and quinolones, are of particular concern. Strains resistant to these last-line antimicrobials are also increasingly reported (4-6), and there are concerns for the potential for widespread emergence of pan-resistant pathogens that are non-responsive to the entirety of our antibiotic arsenal. This problem is exacerbated by market forces that disincentivise investment in research and development of novel classes of antibiotics (7).

It is estimated that 700,000 deaths occur each year as a result of all-cause AMR infections and that by 2050 the number of AMR-attributable deaths annually will increase to 10 million (7). While the health burden of AMR is expected to balloon, the predicted economic impact is also equally staggering, with a global cumulative economic impact of $100 trillion USD anticipated by the mid-point of the century (7). While strategies to combat this phenomenon are increasingly prioritised by governmental agencies and institutions in the animal and human health industries, urgent coordinated and collaborative action is required from professionals across regulatory, agricultural

and clinical environments if we are to be successful in tackling the threat of AMR. This is because the magnitude of the threat posed by AMR is mirrored by its complexity.

2.2 One Health, antimicrobial resistance and infectious diseases

Recently, great efforts have been made to identify the potential sources of AMR. It is well known that antimicrobial selection pressures drive AMR; therefore, sources of AMR evolution and dissemination are likely those which see the highest usage of antimicrobials. Primarily, there are two areas of human activity wherein antimicrobials are used; animal agriculture and in the treatment and prevention of human disease. While it is known that clinical environments play critical roles in the selection for AMR lineages of bacteria, within the United States it is estimated that 70% of antibiotics are used in the raising of livestock (7), highlighting the food-animal industry as another major source of AMR.

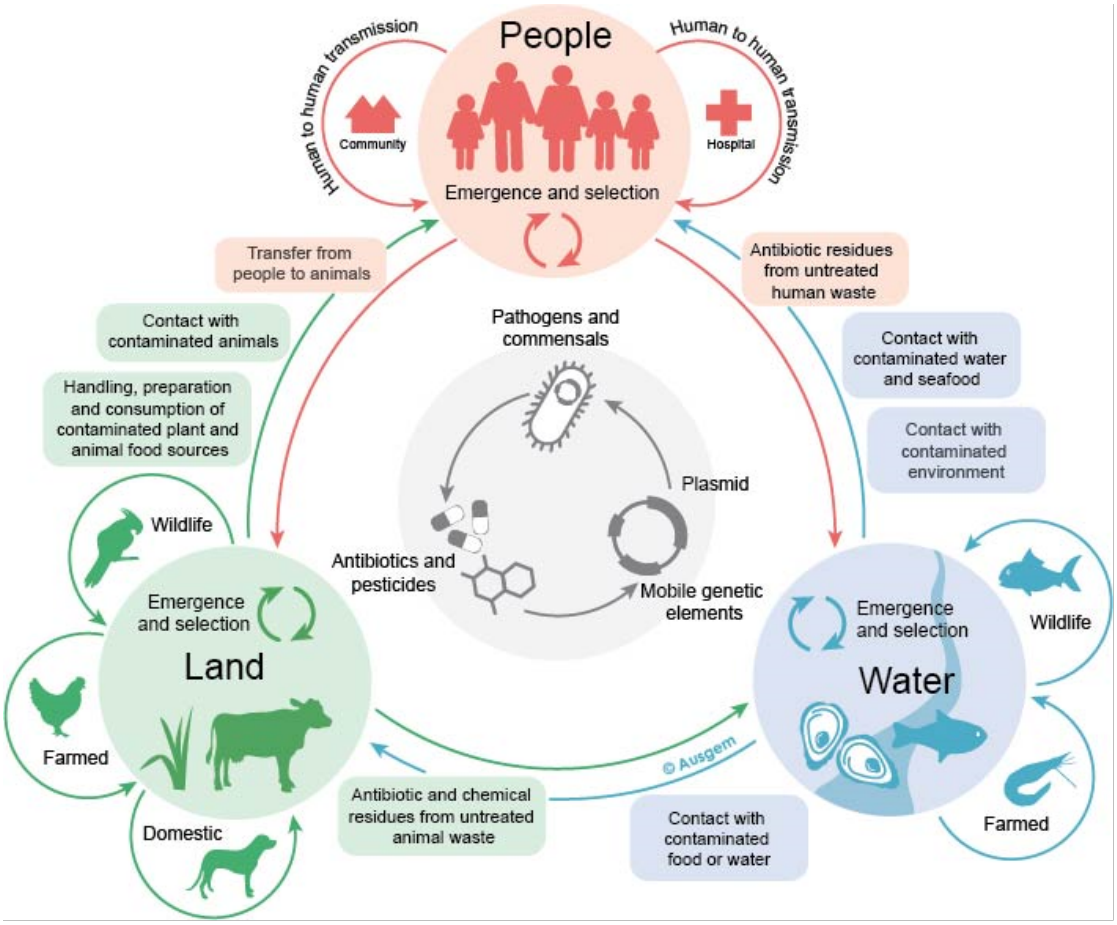

Figure 2.1 – Diagrammatic representation of the mechanisms facilitating movement of microbes, their mobile genetic elements and selective agents like antimicrobials, the basis of One Health ideology

In recent years it is being increasingly understood that microbial populations can transfer to, and establish within, environments other than those from which they originate. Agricultural runoff, human sewage, medical waste and animal-sourced fertilisers constitute just a few examples of the many different vectors facilitating environmental microbial transfer. There are many reservoirs of AMR and pathways that facilitate transfer of AMR bacteria between them, a small number of which are shown in Figure 2.1. However, it is also important to recognise that the spread of AMR is attributable not only to the spread of AMR bacteria.

The movement of genetic material encoding AMR, particularly when present on mobile genetic elements (MGEs), is also a major concern, and will be discussed in greater detail. Additionally, the transfer of antimicrobial compounds and residues between different environments, for example the pollution of natural waterways by antimicrobials of clinical, agricultural or geological origin (e.g. mercury), serves to select for AMR microbes. One Health frameworks are those which consider that these different levels of movement take place between human, agricultural and natural environments. Such approaches are being increasingly adopted by researchers, legislators and other professionals in the study and management of AMR and infectious diseases.

2.3.1 Causes of antimicrobial resistance

The occurrence of AMR is due to the acquisition or mutation of genetic material that facilitates the survival and/or growth of microorganisms in the presence of otherwise efficacious antimicrobial compounds. There are many mechanisms of AMR (Figure 2.2) and the intricacies of the biological architecture that facilitate them are an area of ongoing research. Two primary factors contribute to the emergence of AMR, i) selection and expansion of AMR clones, and ii) lateral gene transfer. The former is more simply explained; AMR selection processes kill off bacterial strains that are susceptible to antimicrobials and those that are resistant survive and propagate. Lateral gene transfer (LGT) is a process whereby genetic material can be exchanged between a donor and recipient cell.

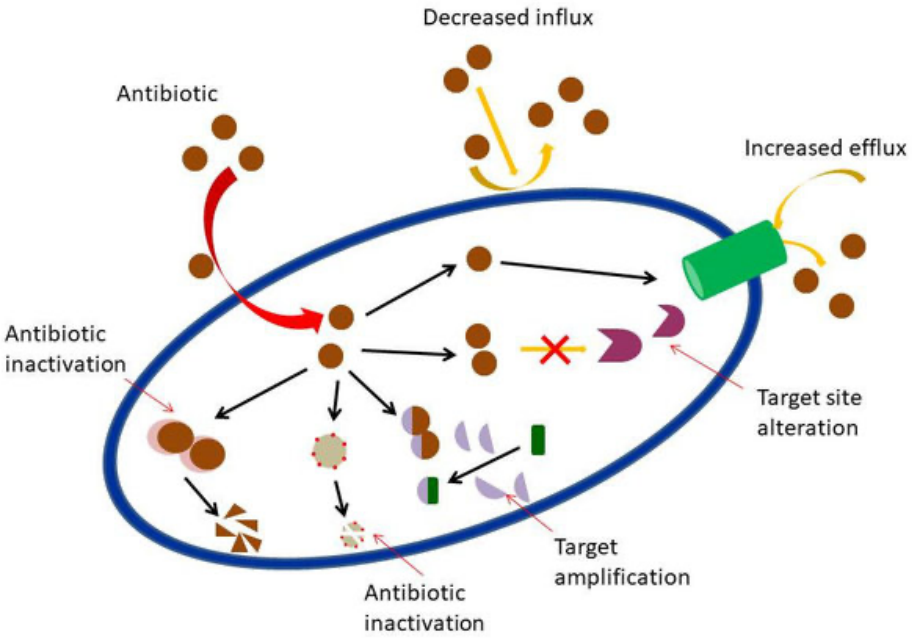

Figure 2.2 – Schematic of some of the mechanisms conferring AMR

2.3.2 Lateral gene transfer

In contrast to clonal expansion, whereby cellular division and genetic replication results in the generation of genetically identical (or near identical) daughter cells, LGT refers to all other processes by which cells can acquire genetic material from non-parent cells. There are several primary mechanisms by which LGT takes place and many types of mobile genetic elements exist through which the process is facilitated (8), as seen in Figure 2.3.

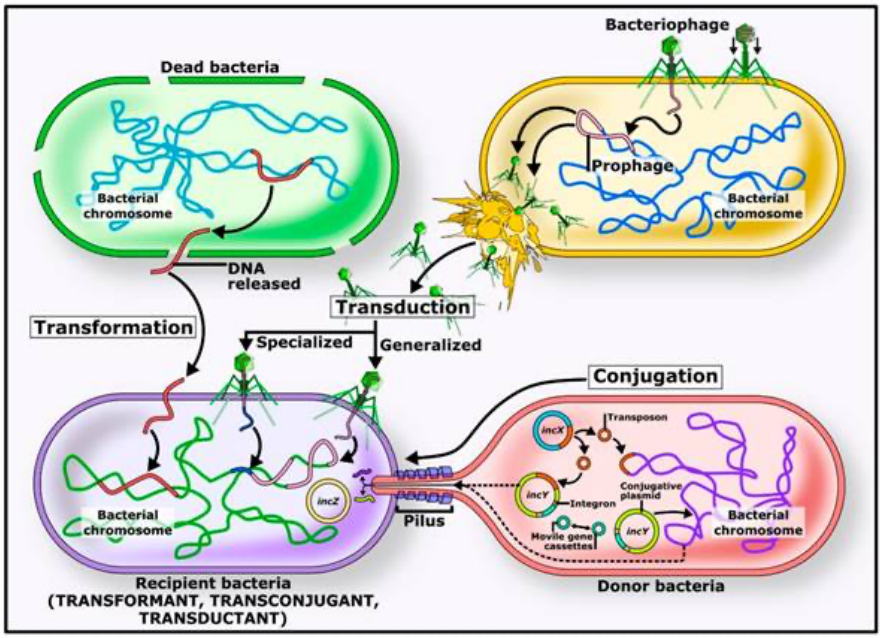

Figure 2.3 – Schematic representation of the primary mechanisms of lateral gene transfer (8)

Transformation involves genetic material being liberated from cells, for example upon cell death, and subsequently encountered and taken up by a neighbouring recipient cell. Secondly, transduction refers to phage-mediated processes of genetic exchange, which play an important role in the spread of AMR and virulence traits. Phages can capture genetic material from the host genome, package it into phage heads and subsequently invade recipient cells and integrate this genetic material into the genomes of their new hosts (9). Thirdly, conjugation, a major contributor to the spread of AMR and virulence, is a process by which plasmids can be transferred through a syringe-like apparatus called a pilus. Plasmids or transposons that contain the genetic material required for pili formation are said to be conjugative, however some non-conjugative plasmids and transposons can hijack the conjugative transfer process and be delivered to a recipient cell *in trans* (10).

2.4 Mobile genetic elements

Many kinds of mobile genetic elements exist which facilitate the processes of LGT. Some of those that are known to have played important roles in the evolution of virulence and AMR in *E. coli* and other bacteria are listed and described below.

2.4.1 Insertion Sequences

Insertion sequences (ISs) are genetic elements that consist of a transposase, and occasionally other coding sequences, between two inverted repeats (IRs). Transposition of an IS element occurs when its transposase, or a related transposase, recognises IRs in a sequence dependent fashion, and occasionally the movement of IS elements results in the movement of adjacent DNA, particularly when this DNA is flanked on both sides by identical or closely related IS elements. A great diversity of IS elements has been reported in the literature, and specific IS elements have been attributed to the dissemination of particular AMR and virulence genes. Among the more notable IS elements are those of the IS*6* family, and IS*26* in particular, which have been attributed to the spread of many AMR genes. For example, a recent study demonstrated IS*26*-mediated transmission of the carbapenem resistance gene $bla_{\text{NDM-1}}$ in a polyclonal multispecies outbreak in a German hospital (11). This lineage of IS elements has also been noted to play important roles in plasmid evolution and the generation and

mobilisation of complex resistance loci (12, 13). The presence of IS elements in multiple copies throughout the genome of a given bacterium is also consequential as it provides additional loci at which homologous recombination events can occur (14, 15).

2.4.2 Plasmids

Plasmids are a diverse group of episomal genetic constructs which can be shared between bacterial strains, in some cases between phylogenetically diverse donor and host cells (16). These MGEs often carry toxin-antitoxin systems to ensure their stable maintenance in host populations, which has important implications for the persistence of their associated genetic cargo, including AMR and virulence genes, within microbial hosts (17). Plasmids can be classified into incompatibility (Inc) groups that reflect their evolutionary history. They are said to be incompatible if cellular division does not result in progenitor cells acquiring both such plasmids, given the presence of both such plasmids in a parent cell (18). To date, twenty eight plasmid Inc groups are recognised (19), however some are thought to have played more important roles in the spread of AMR and virulence than others. Incompatibility types such as IncF, IncHI2, IncX and IncI plasmids are among those that have made the greatest contribution to the spread of AMR throughout Gram-negative bacteria (20) as they frequently harbour AMR genes and elements important in the evolution of AMR such as integrons and IS elements (21-24). In regard to virulence, the plasmid types with the propensity to carry *E. coli* virulence-associated gene cargo are those of Inc type F (25), particularly those of ColV-like virulence plasmids which are discussed in greater detail in Chapter 4.

2.4.3 Clinical class 1 integrons

Integrons are modular genetic constructs which are able to capture and express gene-cassettes, some of which encode an AMR phenotype (26). Multiple cassettes can be present in a single integron, allowing for the accumulation of multiple AMR genes in a single genetic locus (26). This is consequential in regard to AMR evolution because genes that integrate into the host chromosome by other means can interrupt biologically important coding sequences and consequentially result in a fitness cost to the host. Therefore, integrons serve as a means of acquiring genetic material while avoiding deleterious changes to the host genome, a factor contributing to their great success in spreading throughout Gram-negative populations.

Integrons are classified based on the nucleotide sequence of their integrase gene *intI* which encodes a site-specific recombinase responsible for the integration and excision of gene-cassettes. While integrons of class 1, 2 and 3 are observed in *E. coli* and other Gram-negative organisms, the most frequently observed class associated with AMR genes are a subtype of class 1 integrons referred to as the clinical class 1 integron (27). Typical clinical class 1 integrons contain an integrase gene (*intI1*), one or more gene cassettes, a truncated quaternary ammonium compound resistance gene (*qacE*Δ), a dihydropteroate synthase gene which confers resistance to sulphonamides (*sul1*), and an open reading frame (ORF) of unknown function called ORF5 (27). These regions flanking the cassette integration site are referred to as the 5`-CS (*intI1* to *attII*) and the 3`-CS (*qacE*Δ to *sul1*) (Figure 2.4).

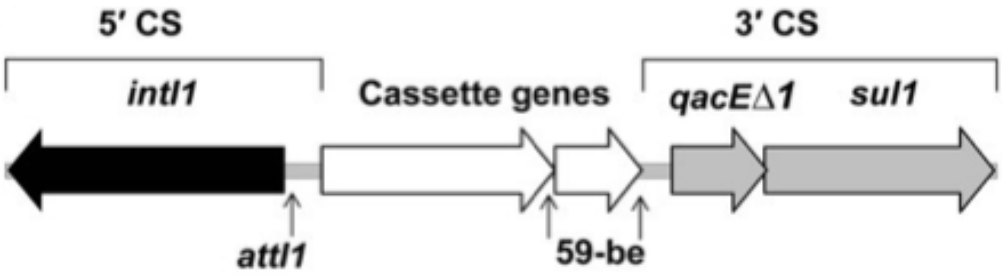

Figure 2.4 – Schematic representation of a class 1 integron, featuring the 5`-CS, a variable cassette region (one cassette shown, however the structure can carry multiple cassettes), and the 3`-CS (187)

Increasingly, atypical clinical class 1 integrons are observed wherein parts of the 5`-CS and/or 3`-CS are missing, typically due to the integration of an IS element, in particular IS*26* (28). These IS*26*-associated integrons are of importance as they not only have enhanced mobility by virtue of their association with IS*26* but they also have a greater capacity to recruit additional IS*26*-mobilised AMR genes, due to IS*26*'s tendency to target and integrate adjacent to other copies of IS*26* (15).

2.3.4 Transposons

Transposons (Tns) are MGEs that encode cellular machinery required for their excision and integration from one genomic context to another. There are two primary classifications of transposons; composite Tns and non-composite Tns. The former are

generated when two identical (or sufficiently similar) IS elements are present proximally to one another in a genomic region. The inverted repeats at either end can be acted on by the IS transposase allowing the excision and mobilisation of the intervening DNA sequence (Figure 2.5). Composite transposons have played important roles in the dissemination of many AMR genes, including Tn*5*, Tn*9* and Tn*6*, which mobilise genes conferring resistance to neomycin/kanamycin, tetracycline and chloramphenicol, respectively (29). A recently identified composite transposon referred to as Tn*6330* has been found to mobilise the colistin resistance gene *mcr-1* through ISApl1-mediated transposition, highlighting the importance of composite Tns in the evolution of AMR (30). The insertion sequence IS*26* is also known to play important roles in the formation of AMR-associated composite transposons (15).

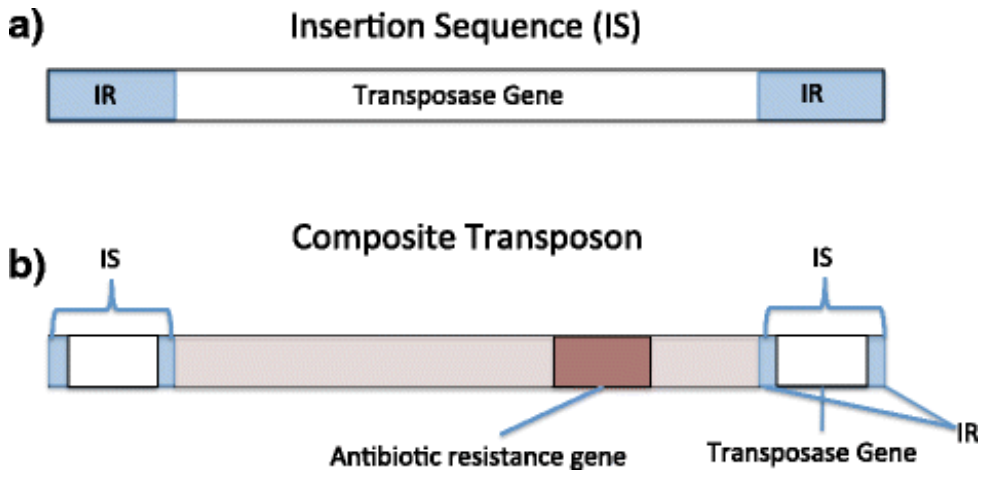

Figure 2.5 – Schematic representation of a) an insertion sequence (IS), flanked by inverted repeats (IRs); b) a composite transposon, flanked by ISs (112)

Conversely, non-composite Tns, also referred to as Tn*3* family elements, encode all genes required for their transposition, including the transposase and resolvase genes *tnpA* and *tnpR* (29). Tn*3* family elements play an important role in the mobilisation and spread of AMR genes throughout bacterial populations, and it is thought that their success in this regard is in part attributable to the association of particular transposons with heavy metal resistance determinants (26). Carriage of heavy metal resistance determinants by Tns is consequential in the spread of these elements as clinical,

agricultural and natural environments are commonly contaminated with heavy metals of anthropological or geological origin, including mercury, zinc and copper (26).

While there are other AMR-associated Tn*3* family transposons within *E. coli* and other Gram-negative species, perhaps the most important is Tn*21*. Tn*21* and other related Tns carry genes required for their transposition (*tnp)*, a class 1 integron (typically carrying one or more AMR conferring gene cassettes as well as a *sul1* gene in the 3`-CS) and the heavy metal resistance operon *mer* (26). It is thought that heavy metal contamination in environmental and agricultural settings has played important roles in the spread of integrons and their associated AMR gene cargo (31). Transposons of the Tn*21* family are commonly found within AMR bacteria from a variety of different sources and in a variety of genomic contexts (26), and it is important to note that gene-cassettes, integrons, Tns and plasmids are often nested within one another (32), as visualised in Figure 2.6.

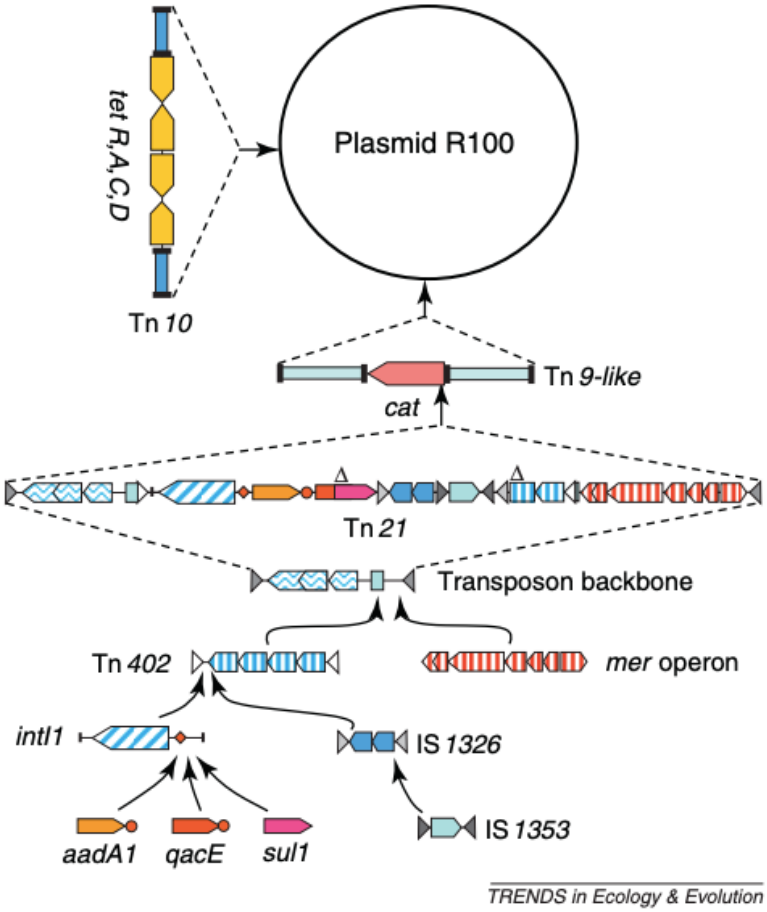

Figure 2.6 – A schematic representation of the archetypal composite and nested structure of the multidrug resistance plasmid R10 (32)

2.3.5 Integrative mobilisable elements and *Salmonella* genomic island 1

Integrative mobilizable elements (IMEs) are a diverse, widespread and somewhat poorly understood class of MGEs that are thought to play important roles in lateral gene transfer. These elements carry genes encoding cellular machinery for their own excision and integration and hijack the conjugative machinery of other co-resident MGEs to facilitate IME spread via LGT (33). Many IMEs also carry AMR genes and some carry multiple, with the most clinically significant AMR-associated IME being those of *Salmonella* genomic island 1 (SGI1) and SGI1-related elements (SGI1-REs) (33).

Salmonella genomic island 1 was first identified in the epidemic multidrug resistant lineage of *Salmonella enterica* Typhimurium DT104 (*S.* Typhimurium DT104) (34). Constituting a 43 kb IME, SGI1 houses a complex class 1 integron conferring resistance to five different classes of antimicrobials (beta-lactams, amphenicols, aminoglycosides, sulphonamides and tetracyclines). Strains of *S.* Typhimurium DT104 were first isolated from wild birds in the 1980s, subsequently became epidemic in cattle in Britain and also common within other food-animal operations (35). Outbreaks of *S.* Typhimurium DT104 infection in both humans and food-animals have been documented internationally and SGI1 and SGI1-REs have now been reported in other serovars of *Salmonella enterica*, other species of *Salmonella* and subsequently in other genera, including other human pathogens such as *Proteus*, *Enterobacter*, *Morganella, Acinetobacter* and *Providencia* (35-39).

Evidence also indicates that SGI1 has the potential to integrate into many other genera, some of which constitute major human pathogens, including the clinically important ESKAPE pathogens *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*, by virtue of SGI1-REs integration site being part of the highly conserved GTPase *trmE* (40, 41). This IME, and elements related to it, are increasingly appearing within clinical and agricultural environments and some recent variants have been reported to have acquired genes conferring resistance to critically important antimicrobials (CIAs) used as last-line treatments for infectious diseases, such as $bla_{\text{NDM-1}}$, $bla_{\text{VEB-6}}$ and $bla_{\text{CTX-M-15}}$ (42-44). Notably, IS*26* is proximal to these CIA resistance conferring genes in two of these three examples, and reported to cause rearrangements in SGI1-RE structures more

broadly (45), highlighting the important role of IS elements such as IS*26* in the evolution of SGI1-REs.

Evidence has indicated that SGI1 is stably maintained in the host chromosome for 351 successive generations even in absence of selection pressure (46), and it has been proposed that this characteristic is likely shared by other IMEs including SGI1-REs (33). This has important implications as it demonstrates that once an SGI1-RE is acquired by a particular lineage it can be expected to persist chromosomally, evolve, and potentially expand via processes of vertical and lateral gene transfer. The spread of SGI1-related elements and other IMEs is potentially of major clinical significance and an area requiring further investigation given the evidence surrounding their broad host range.

2.4 *Escherichia coli* as a One Health organism

E. coli is found in the gastrointestinal tract of most vertebrates as well as in terrestrial and aquatic environments; natural and man-made (47). It is produced *en-masse* in food-animal production systems wherein these bacteria are routinely exposed to antimicrobials including antibiotics and heavy metals, which have historically been used as food additives to prevent the occurrence of animal disease and as a nutritional supplement, respectively (31). While the usage of antibiotics for growth promotion is banned within Australia and many countries abroad, some studies suggest that AMR populations of bacteria may persist even in the absence of continued exposure to antibiotics (48). This arises because compensatory mutations that ameliorate the fitness impost posed by carrying MGE enable bacteria to carry them (48).

Food-animal associated *E. coli* are frequently MDR, particularly in countries where regulations regarding antimicrobial usage are lax and stewardship practices are poorly enforced (49, 50). AMR and potentially pathogenic *E. coli* have been isolated from animal food-products and outbreaks of infectious disease caused by *E. coli* have been attributed to the consumption of meat, eggs and dairy (47). However, due to fertilisation of food crops with food-animal litter being commonplace, fresh produce has been demonstrated to be contaminated with virulent and AMR *E. coli* (51). Shiga-toxin producing *E. coli* (STEC) outbreaks in many countries have been linked to the consumption of contaminated vegetables, for example, with a single outbreak of STEC

in the United States traced to contaminated romaine lettuce causing 167 cases of illness and 85 hospitalisations across 27 states (52).

Similarly, *E. coli* inhabit the gastrointestinal tracts of humans, and therefore clinical and community environments are commonly contaminated with these bacteria. Both raw and treated sewage have been shown to have higher prevalence of AMR bacteria than normal surface waters, and while wastewater treatment plants are known to be effective in reducing the overall load of AMR bacteria within sewage effluent they have been noted as hotspots wherein AMR may emerge (53, 54). Similarly, agricultural runoff has been shown to contaminate waterways with *E. coli*, MGEs and antimicrobials (55-57). Particular lineages of *E. coli* have been shown to persist within environmental reservoirs, highlighting that evolution of AMR traits continues to take place in natural environments (58).

The flow of *E. coli* and AMR-associated MGEs between natural, agricultural and clinical environments is of concern as it is thought that the pangenome of environmental bacteria, particularly those associated with the soil microbiome, play an important role as a source of AMR conferring genes and also other genetic elements involved in AMR evolution (59). Here via lateral gene transfer they can acquire novel AMR genes from the environmental resistome, a process hastened by the presence of antimicrobial contaminants that are common in environmental settings (31). Subsequently they can be mobilised by the movement of water, migratory animals and other forces.

It is estimated that by 2050 AMR *E. coli* will cause approximately 3 million human deaths per annum and be responsible for a cumulative economic impact of approximately $40 trillion USD (7). Their ubiquity, combined with their role as a major human and animal pathogen and their propensity to acquire and disseminate AMR-associated MGEs, makes *E. coli* an ideal organism to study within a One Health framework.

2.4.1 Phylogenetic classifications of *E. coli*

Many approaches have been developed for the phylogenetic classification of organisms in all kingdoms of life, including *E. coli*. Such classifications facilitate the comparisons of members of the species, as certain lineages of *E. coli* have a tendency to be associated with particular niches and differ in their capacity to cause various pathologies (47). Prior

to the discussion of the nuances of *E. coli* pathotypes and niche-associations, some of the phylogenetic classification schemes used to classify *E. coli* will be outlined.

2.4.2 Serotyping

Serotyping of bacteria involves characterisation based on the presence of particular cell surface antigens. Traditionally performed *in vitro* using antisera, serotype determinations can now be performed *in silico* using WGS data. In *E. coli*, serotyping can be used to classify the O (lipopolysaccharide; usually present), H (flagellar; usually present) and K (capsule; variably present) antigens. Currently there are upwards of 186 known O-types, at least 53 known H-types (60) and 80 or more known K types (61), however K-typing is less commonly performed than O- and H-typing. The most well-known serotype of *E. coli* is O157:H7, a lineage commonly attributed to STEC infections that can result in bloody diarrhoea and haemolytic uremic syndrome (47). Notably, not all O157:H7 strains cause STEC infections, and not all STEC infections are caused by O157:H7 strains (47). Serotype-pathology associations are therefore non-absolute. STEC are prevalent in the faeces of cattle and sheep and other ruminant species in Australia and internationally (62-64). It should also be noted that there is some capacity for serotype to be transferred laterally, particularly in the case of O-type genes (65).

2.4.3 Phylogrouping

Phylogrouping of is a scheme of phylogenetic classification of *E. coli* that is based on the presence or absence of genes and gene fragments in a given strain. While the original scheme consisted of four phylogroups, A, B1, B2 and D (66), there have been several modifications to the scheme whereby additional groups have been added (67). This thesis employed the original Clermont scheme with just four classifications as it is widely utilised; please note that all subsequent mentions of phylogroups within this thesis are referred to in the context of this particular scheme.

Traditionally, as with serotyping, phylogrouping was performed using a polymerase chain reaction in combination with gel electrophoresis, based on the presence or absence of amplicons of the conserved genes *chuA* and *yjaA* and the conserved nucleotide sequence referred to as TspE4.C2 (66). However, phylogroup determination can now also be performed computationally. The literature suggests that phylogroups A

and B1 are primarily associated with commensal strains and IPEC, while phylogroups B2, and to a lesser extent D, are more commonly attributed to ExPEC infections (68).

2.4.4 Multilocus sequence typing

Multilocus sequence typing (MLST) is a moderately high-resolution scheme of phylogenetic classification in which strains are grouped into sequence types (STs) based on sequence variation in seven highly conserved genes. Sequence types that exhibit a difference in one of these alleles are said to be of the same clonal complex (CC). Some STs have been shown to have associations with pathological and environmental niches. In some cases, the fitness traits of a given ST that make it successful in a particular environment are better understood. For example, research has indicated that the success of ST131, a pandemic ST associated with multidrug resistant ExPEC infections, may be attributable to several primary genetic underpinnings. These include reduced uptake of antimicrobials through OmpC-mediated porin downregulation (69), enhanced biofilm formation compared with non-ST131 strains (70), and resistance to fluoroquinolones and extended spectrum beta-lactamase through acquisition of single nucleotide polymorphisms (in *gyrA* and *parC*) and the ESBL gene $bla_{\text{CTX-M-15}}$ (71). While the reasons for the global dominance of ST131 as an epidemic ExPEC lineage are not completely understood, research has shed light on some factors potentially involved. For other STs with particular niche associations, the reasons for this success are yet to be explored. Such is the case for ST117 and ST95, and their strong association with poultry (72-74).

2.4.5 Core-genome multi-locus sequence typing

While the concept of core-genome multi-locus sequence typing (cgMLST) is similar to that of MLST, the former provides much greater phylogenetic resolution than traditional MLST approaches as it defines sequence types based on 2513 conserved genes, rather than the seven genes used in MLST (75). Similarly, with closely related MLSTs being clustered into clonal complexes, Hierarchical Clustering of cgMLST (HierCC) is a system by which strains can be clustered into groups of greater or lesser relatedness. A hierarchical cluster (HC) of 0 indicates that strains within this cluster carry identical

alleles across all core genes within the cgMLST scheme for this species, while an HC of 50 indicates that no more than fifty alleles in this cluster vary between strains.

2.5 Pathotypes of *E. coli* – varieties and genetic underpinnings

There are three major phenotypic classifications of *E. coli*, including commensal strains, intestinal pathogenic *E. col* (IPEC) and Extraintestinal pathogenic *E. coli* (ExPEC). Pangenomic modelling has indicated that the *E. coli* pangenome may consist of more than 18,000 families of orthologous genes; genomic diversity reflective of the spectrum of ecological niches this species inhabits (76). It is thought that the accessory genome of *E. coli* determines its ability to occupy particular pathogenic niches, and that carriage of particular virulence-associated genes indicate the ability of a given strain to cause disease. For example, intestinal pathogenic *E. coli* lineages can typically be designated as such genotypically by the presence of particular toxin genes (47).

However, differentiating commensal and ExPEC strains is more difficult, with no single genes or gene combinations necessarily determinant of ExPEC status (77). This is particularly true for human ExPEC, yet there have been some successful attempts made to identify the minimal genetic predictors of a subset of ExPEC that infect poultry animals (78). This latter group are referred to as avian pathogenic *E. coli* (APEC) and will be discussed in greater detail. Given that hybrid pathotypes, strains that carry hallmarks of both IPEC and ExPEC lineages, are increasingly reported in the literature, the shortcomings of genotypic classification of *E. coli* should be noted.

2.5.1 Intestinal pathogenic *E. coli*

Intestinal Pathogenic *E. coli* (IPEC) are a major cause of gastrointestinal illness within both humans and animals and are typically associated with diarrhoeal diseases. Many different varieties of IPEC exist, however shiga-toxigenic *E. coli* (STEC) are perhaps the most clinically significant. While STEC are phylogenetically diverse there are a cluster of major STEC serotypes O157:H7 (ST11) and six non-O157 STEC including O111, O103, O145, O26, O121 and O45, which cause diarrhoeal illness and in severe cases multiple organ failure (64). These STEC are considered part of the normal flora in the hindgut of ruminants and are frequently shed in faeces (47). Within the agricultural sector, IPEC lineages also cause severe disease burden, particularly within piggeries where they

cause post-weaning diarrhoea in piglets which causes significant reductions in the productivity of these operations (79).

2.5.2 Extraintestinal pathogenic *E. coli*

Aside from gastrointestinal illnesses, many lineages of *E. coli* are capable of causing extraintestinal infections and are therefore known as Extraintestinal Pathogenic *E. coli* (ExPEC). Evidence indicates that the primary reservoir of human ExPEC is in the human gut (77). Among human populations, ExPEC cause urinary tract infections (UTIs), sepsis or blood stream infection (BSI) and neonatal meningitis, however a diverse array of soft tissue infections and other pathologies are also commonly attributed to *E. coli* (77). While there is not yet sufficient evidence to suggest that these groups are necessarily distinct from one another at a genotypic or phylogenetic level, the terms uropathogenic *E. coli* (UPEC), Human sepsis-associated *E. coli* (SEPEC) and Neonatal meningitis-associated *E. coli* (NMEC) are used to describe strains attributed to cases of UTI, sepsis and neonatal meningitis, respectively.

Urinary tract infections are the most common community-acquired infection caused by Gram-negative bacteria, the leading cause of bacteraemia in adults and the second most prevalent cause of neonatal meningitis (80). One in two women will contract a UTI in their lifetime, with female, immunocompromised, elderly and catheterised individuals at greater risk (81). While the prevalence of *E. coli* as causative agents of UTIs varies between studies, the species is thought to be the most prevalent cause of such infections (82, 83). Urinary tract infections have been found to be increasingly resistant to antimicrobials (84). In the absence of effective treatment ExPEC can ascend the urinary tract, access the kidneys causing pyelonephritis and subsequently enter the blood, a process known as urosepsis (47).

Cases of sepsis require urgent treatment as septic shock and death can result in a matter of hours or days, with the disease burden of the condition reflected by its rank as the single most costly healthcare problem in the United States (85). It is important to note that sepsis can occur via means of infection other than urosepsis, with other common sources of infection including bacterial contamination of the blood following penetrative traumas, catheterisation or invasive surgery. Regardless, the blood is a hostile environment for bacteria and colonisation of the blood by *E. coli* is thought to be

mediated in part by the production of protectins such as those encoded by *iss* (increased serum survival) and *ompT* (outer membrane protein) (86, 87).

Free iron is severely limited in extraintestinal sites to prevent the growth of pathogens therein, and iron acquisition serves as a major challenge to the growth of *E. coli* in such sites. It is therefore unsurprising that ExPEC frequently carry genes involved in iron acquisition pathways, including aerobactin and yersiniabactin operons *iro* and *iut*/*iuC*, respectively (74, 88, 89). Similarly, factors such as carriage of *ibeA* (encoding invasion of brain endothelium) and capsular antigens have been shown to be characteristic of NMEC and shown to have a greater capacity to form biofilms than those isolated from the guts of healthy individuals (90-92). Notably, there is evidence that NMEC, UPEC and SEPEC share genotypic and evolutionary relatedness with other members of the species – those associated with poultry (91, 92).

2.5.3 Avian pathogenic *E. coli*

Avian pathogenic *E. coli* are a subtype of ExPEC which are associated with extraintestinal infection in avian hosts, especially poultry. Manifestations of APEC pathologies are diverse and the mechanisms of pathogenesis are numerous, but are broadly referred to as colibacillosis (93). Primarily, APEC are attributed to systemic infections of poultry following the inhalation of airborne faecal particles wherein a subset of this faecal flora has a capacity to survive and reproduce, particularly in immunocompromised birds afflicted with other viral respiratory pathogens (93). Invasins facilitate epithelial cell invasion compromising the integrity of the host lung tissue and subsequently allowing access to the cardiovasculature, with the resultant sepsis often proving fatal or warranting animal culling (93). Colibacillosis is thought to be the most important bacterial disease of poultry and results in extensive condemnation of broiler carcasses, reduced productivity of laying hens and is frequently attributed to the death of chicks (94).

The genomic underpinnings of the APEC phenotype are relatively well studied and it is generally agreed upon that a defining characteristic of APEC is the carriage of large virulence plasmids of incompatibility group F (78, 95). These plasmids are typically categorised based on their carriage of colicins, with ColV plasmids and ColBM plasmids being among the most common types identified within APEC lineages (96). These

plasmids typically host extensive virulence associated gene cargo that plays a critical role in pathogenesis. This cargo includes iron-acquisition genes (*iut*/*iuc*, *iro*, *ets*, *eit*, *sit*, *irp2*), protectins (*iss*, *traT*) and the toxin chaperone *hlyF* (Figure 2.7) (25, 95). Some of these genes also occur in chromosomal genomic contexts in association with pathogenicity islands, however, and there is evidence of strains of *E. coli* capable of causing extraintestinal infections in poultry which do not appear to carry these large virulence plasmids (78, 86).

Research has indicated that many APEC VAGs are also present in human ExPEC strains, leading to speculation that APEC may be zoonotic pathogens (73, 97, 98). This conjecture is supported by the observation that there is also phylogenetic overlap between *E. coli* causing human and avian extraintestinal infections, as is the case with sequence types ST117 and ST95 (73, 74, 99, 100).

Furthermore, human ExPEC frequently carry many of the same iron-acquisition genes and protectins, and some human ExPEC that carry ColV plasmids have also been identified (91, 101, 102). This data suggests that APEC may not only constitute zoonotic

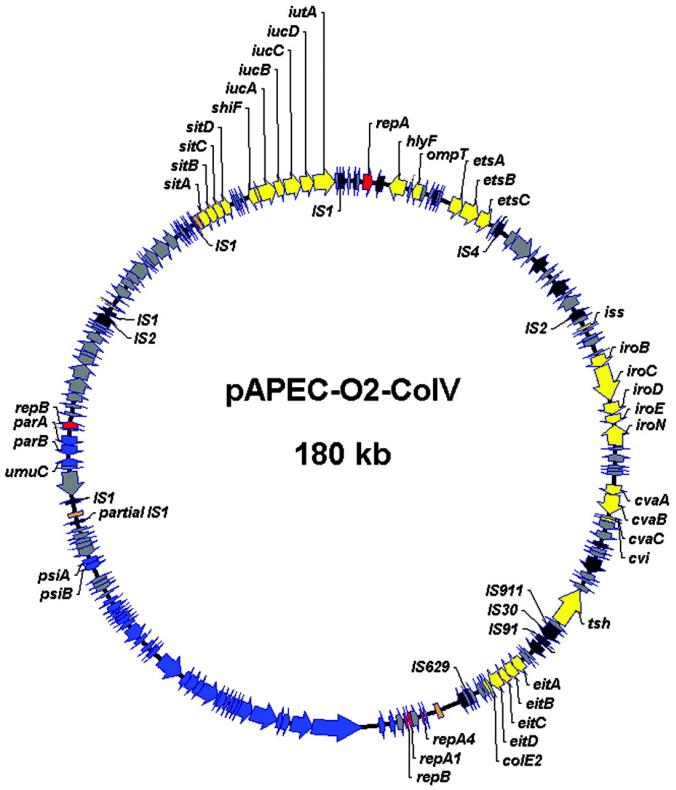

Figure 2.7 – Schematic representation of a large IncF ColV plasmid (122). Yellow – virulence genes; blue - plasmid transfer and maintenance genes; red – plasmid replication genes; grey – genes of unknown function; black - mobile genetic elements

potential but may also have a capacity to serve as donors of plasmids which can enhance the virulence potential of otherwise commensal *E. coli*. A recent study reported on a commensally-associated lineage of *E. coli* ST58 causing urosepsis which was found to carry a ColV virulence plasmid, with the authors hypothesising this plasmid acquisition event may have played an important role in conferring its ExPEC potential (101). While this is speculative, experimental evidence lends credit to this hypothesis, with a recent study demonstrating ColV plasmid acquisition by a commensal *E. coli* enhanced its growth in human urine, human complement resistance and colonisation of the murine kidney (103). Furthermore, acquisition of ColV by *Salmonella enterica* serovar Kentucky was considered a defining event during recent outbreaks of disease caused by this organism (104).

Additional insight into the APEC and their associated MGEs is an area requiring urgent investigation, and genomic data on Australian APEC is particularly lacking. Garnering a greater understanding of the movement of bacteria between poultry reservoirs and those associated with other human, agricultural and natural environments is an important part of developing One Health models of infectious diseases and AMR. This is particularly urgent given that poultry production is set to double between 2005 and 2050 (105), and while the growth of this sector will likely result in increased cost of APEC infections for the poultry sector it is important that other potential threats can be identified and if necessary, mitigated.

Chapter 3: Methodological Overview

3.1 Study Design

Please note that much of the information in the following sections relate primarily to Chapter 4 and 5 of this thesis, as these chapters were developed using sequence data generated in-house, whilst the sequence data utilised in Chapter 6 was sourced from publicly available genome sequence databases.

3.1.1 APEC collection under investigation

The primary collection under study consisted of 95 strains of APEC originating from at least twelve different commercial poultry operations from around Australia (Victoria, New South Wales, Queensland and Western Australia). The collection spans a nine-year period between 2007 and 2015 and was sourced primarily from broiler and layer chickens with some strains also sourced from turkeys and ducks.

Only *intI1*-positive strains were chosen for whole genome sequencing and analysis, as *intI1* has been shown to be a reliable proxy for multidrug resistance (31). These strains were sourced from a larger collection (n=256) of shared origin, of which 50% of strains (n=123/256) were found to carry a class 1 integrase (*intI1*) by PCR. Following DNA isolation, preparation, sequencing, quality control and removal of contaminated and duplicate strains, 95 strains remained for *in-silico* analysis.

3.1.2 Whole genome sequencing as a methodology for genomic epidemiology

The widespread deployment of genomic epidemiological investigations is a critical undertaking in the fight against AMR and infectious diseases. It is only through a greater understanding of the interplay between human, animal and environmental health, in regard to AMR and infectious agents, that strategies can be designed to mitigate these threats.

Whole genome sequencing is a suitable approach to this end. It has many advantages over more traditional molecular biology techniques such as polymerase chain reactions (PCRs). Firstly, the identification of novel emerging AMR genes and MGEs responsible for the dissemination is an ongoing process. Utilisation of WGS allows for rapid back

screening of genomic datasets for novel DNA sequences of interest. Additionally, WGS is highly scalable. While multiplex PCRs can be designed for the screening for several genes in a single assay, it is commonplace to screen thousands of WGS datasets for thousands of genes of interest (75). For example, the tool BIGSI allows for the screening of 455,632 bacterial genomic datasets for a sequence of interest, and return within less than a minute the strains in which this sequence has been identified as well as their probable genus and species (106). High resolution phylogenetic analysis are also facilitated by analysis of WGS data and increasingly relied upon to provide insight into the epidemiology of disease outbreaks (107-109).

While reproducibility of research has been highlighted as an issue across scientific domains, the reproducibility of *in silico* analyses has also received a lot of scrutiny (110, 111). Usage of different versions of software and databases can lead to discrepancies in results in the output of analytical pipelines. Similarly, manual curation of large datasets is inherently irreproducible in absence of detailed curation of all changes made and the justification for them. Worsening matters, scrutiny of metadata associated with strains uploaded in public databases reveals the sources of many strains are misappropriated, compromising the trustworthiness of analysis attempting to resolve the associations of particular microbial communities with environmental settings. Significant advancements have been made in recent years in regard to development of software that facilitates reproducibility in bioinformatic analyses (111), however, and public WGS databases still serve as powerful tools in genomic epidemiological investigations. It is, after all, the collaborative power of WGS data which is perhaps its greatest advantage over traditional molecular approaches to epidemiological investigations and AMR surveillance.

3.2 Methodological Overview

Note: As comprehensive methodologies are provided within each results chapter, the following is a brief overview of the methodology utilised in this thesis.

3.2.1 DNA isolation, whole genome sequencing and genome assembly

Glycerol stocks previously generated from pure cultures of APEC were used to inoculate LB broth. Following overnight growth, the cells were pelleted and lysed, and the

liberated DNA was then extracted, sequenced and assembled via different approaches for short- and long-read sequencing data.

Short-read – Illumina platform

Short-read WGS using Illumina requires relatively little DNA (~1 nanogram) and therefore a brief protocol involving an off-the-shelf kit is suitable for extraction of the required quantity and quality of DNA. This project utilised the Bioline^{TM} ISOLATE II Genomic DNA Kit for this purpose. Short-read sequencing was performed using an Illumina HiSeq 2500 and genomes were assembled (*de novo*). Chapter 4 utilised the assembly pipeline A5, however in Chapter 6 short-read sequence data was assembled using Shovill (a SPADES assembly pipeline) as it is more widely utilised by researchers due to its performance and ease of use. Short-read WGS and assembly typically results in genomes with relatively low contiguity, preventing the linkage of particular genomic loci with chromosomal or episomal contexts (112). To circumvent this issue, long-read sequencing was employed.

Long-read - Oxford Nanopore platform

Long-read sequencing using the Oxford Nanopore platform requires several micrograms of concentrated, pure and high molecular weight DNA. Therefore, DNA for use in this latter application was isolated using a slightly modified phenol-chloroform extraction, involving xanthogenate-SDS buffer to lyse cells (113) and a subsequent bead-based clean up. Long-read sequencing was performed using an Oxford Nanopore MinION, whilst hybrid (*de novo*) assemblies, used to generate more complete assemblies for more detailed analysis of particular genomic structures of interest (Chapter 5), combined short-read and long-read sequence data through the use of the tool Unicycler. Unicycler is a hybrid assembler which performs a short-read assembly, subsequently bridges assembly gaps through alignment of long-reads and performs a polishing step whereby short-reads are reemployed to improve the accuracy of the base-calls across the assembly.

3.2.2 Phylogenetic analysis

Multilocus sequence typing (MLST)

Multilocus sequence typing (MLST) allows classification of *E. coli* into sequence types (STs) based on sequence variation within seven conserved genes. Carriage of particular alleles of these seven genes results in an allelic profile corresponding to a specific ST. While another scheme of sequence typing exists for *E. coli,* it is less widely utilised than the Achtman scheme (114), which was utilised throughout this thesis.

Phylosift

Phylosift is a tool originally designed for the analysis of metagenomic sequence data but has been demonstrated to be a robust tool for inferring the phylogenetic relatedness of diverse strains of *E. coli* (Chapter 4). It is useful as its ability to approximate the phylogenetic distances between different strains under analysis is independent of the biases associated with reference based phylogenetic analyses and also those encountered when analysing particularly large and diverse collections of bacteria (there is an inverse relationship between the phylogenetic diversity of a collection and the size of the core genome shared by these organisms). Evolutionary distance between isolates is calculated based on differences in the amino-acid sequences corresponding to the coding sequences of thirty-seven highly conserved genetic loci.

Single-nucleotide polymorphism based phylogenetic analysis

While Phylosift is useful for determining the phylogenetic relatedness of diverse strains of *E. coli*, it is of too low phylogenetic resolution to allow the comparison of closely related strains that are of the same sequence type or, especially, potential clonal outbreaks. For such comparisons, SNP analysis is often much more informative and can allow quantification and location of the SNPs that strains share or those by which strains differ. While several tools exist for performing SNP analysis on bacterial genomes, Snippy was utilised due to its widespread usage within bacterial phylogenetic analysis. As *E. coli* genomes are known to be frequently subjected to recombination events, recombination filtering using Gubbins serves to filter out particularly high density regions of SNPs that are usually artefacts of homologous recombination events rather than SNPs that serve as proxies of evolutionary divergence that have accumulated as a

result of random mutations. The phylogenetic trees used to generate Figures 4.2 and 4.3 and 6.2, respectively, were generated using Snippy and Gubbins.

Generation and visualisation of phylogenetic trees using FastTree2, ggtree and the Interactive Tree of Life (iTOL)

FastTree2 was utilised to generate phylogenetic trees from the alignment files generated by Phylosift and Snippy/Gubbins using a generalised time-reversible model. These tree files were then visualised using ggtree (Figures 4.2, 4.3 and 6.2) or iTOL (Figure 4.1), the former is an R package used for the visualisation and annotation of phylogenetic trees and the latter is a web-based graphical-user-interface tool with the same purpose. Both tools were utilised as iTOL lends itself better to trees in circular format than ggtree does, however ggtree generates figures in a more customisable and more easily reproducible fashion than iTOL.

Genotypic analysis

Genomic datasets generated by short-read sequencing were interrogated using the tool Antibiotic Resistance Identification by Assembly (ARIBA) for their carriage of genes associated with virulence, AMR, plasmid carriage, phylogroup classification and serotypes as well as other genes of interest. The specific versions of these databases are available at https://github.com/CJREID/custom_DBs.

Genomic annotation and schematisation

Annotation of SGI1-related elements in Chapters 5 and 6 were performed using a combination of BLASTn (command-line) and MegaBLAST (available at https://blast.ncbi.nlm.nih.gov/). This analysis was used to identify and locate the precise genomic coordinates of genetic loci associated with previously published variants of SGI1-REs. Where the identities of particular genetic loci could not be resolved through comparative analysis with published SGI1-RE sequences, these loci were screened against the NCBI nucleotide database using MegaBLAST or against the ISfinder database (https://isfinder.biotoul.fr/blast.php).

Once the identities of these genetic loci were determined, the schematics of the structures, as well as their associated GenBank files used for uploading these sequences

to sequence repositories, were generated using the proprietary software Snapgene (GSL Biotech).

In cases where nucleotide identity maps were required to visualise the relatedness of SGI1-REs (such as in Figure 6.1 and 6.3), EasyFig was used to compare the GenBank files generated through manual annotation in Snapgene. Subsequent colouring and other cosmetic changes were performed using the image editing software Adobe Photoshop to facilitate the generation of publication quality figures.

Chapter 4: Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene

Authors

Max L. Cummins[1], Cameron J. Reid[1], Piklu Roy Chowdhury[1], Rhys N. Bushell[2], Nicolas Esbert[2], Kelly A. Tivendale[2], Amir H. Noormohammadi[2], Shaiful Islam[2], Marc S. Marenda[2], Glenn F. Browning[2], Philip F. Markham[2] and Steven P. Djordjevic[1*]

Author affiliations

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

[2] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, and Werribee, Victoria 3030, Australia.

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Published in Microbial Genomics, 23 January 2019. https://doi.org/10.1099/mgen.0.000250.

4.1 Declaration

M.L.C. was responsible for project administration, conceptualisation, and investigation, writing and execution of software, formal curation, analysis, validation and visualisation of data and writing the original manuscript draft and editing.

C.J.R, P.R.C: responsible for reviewing and editing draft works.

R.N.B, N.E, K.A.T, A.H.N, S.I, M.S.M, G.F.B: responsible for sample collection and metadata collection.

S.P.D. was responsible for supervision, resource provision, formal analysis, reviewing and editing draft versions.

All authors have read and agreed to the published version of the manuscript.

4.2 Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene

This publication addresses Aims 1 and 2 of the thesis. Please note that figure numbers have been changed from the original publication to ensure they are consistent with the thesis. Some of these figures in this thesis are best viewed via the attached files.

4.2.1 Abstract

Avian pathogenic *Escherichia coli* (APEC) cause widespread economic losses in poultry production and are potential zoonotic pathogens. Genome sequences of 95 APEC from commercial poultry operations in four Australian states that carried the class 1 integrase gene *intI1*, a proxy for multiple drug resistance (MDR), were characterized. Sequence types ST117 (22/95), ST350 (10/95), ST429 and ST57 (each 9/95), ST95 (8/95) and ST973 (7/95) dominated, while 24 STs were represented by one or two strains. FII and FIB *repA* genes were the predominant (each 93/95, 98 %) plasmid incompatibility groups identified, but those of B/O/K/Z (25/95, 26 %) and I1 (24/95, 25 %) were also identified frequently. Virulence-associated genes (VAGs) carried by ColV and ColBM virulence plasmids, including those encoding protectins [*iss* (91/95, 96 %), *ompT* (91/95, 96 %) and *traT* (90/95, 95 %)], iron-acquisition systems [*sitA* (88/95, 93 %), *etsA* (87/95, 92 %), *iroN* (84/95, 89 %) and *iucD*/*iutA* (84/95, 89 %)] and the putative avian haemolysin *hylF* (91/95, 96 %), featured prominently. Notably, mobile resistance genes conferring resistance to fluoroquinolones, colistin, extended-spectrum β-lactams (ESBLs) and carbapenems were not detected in the genomes of these 95 APEC but carriage of the sulphonamide resistance gene, *sul1* (59/95, 63 %), the trimethoprim resistance gene cassettes *dfrA5* (48/95, 50 %) and *dfrA1* (25/95, 27 %), the tetracycline resistance determinant *tet(A)* (51/95, 55 %) and the ampicillin resistance genes $bla_{\text{TEM-1A/B/C}}$ (48/95, 52 %) was common. IS*26* (77/95, 81 %), an insertion element known to capture and mobilize a wide spectrum of antimicrobial resistance genes, was also frequently identified. These studies provide a baseline snapshot of drug-resistant APEC in Australia and their role in the carriage of ColV-like virulence plasmids.

4.2.2 Data summary

1. Ninety-five pairs of short-read data of avian pathogenic *E. coli* sequenced for this study have been deposited in the NCBI Short Read Archive under Study ID 479542. Additionally, draft genome assemblies have also been uploaded and are accessible under this same study

ID. Individual sample accession numbers can be found in Table S1 (Additional file 4.1) (also available at the online version of this article).

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA479542

4.2.3 Impact Statement

Avian pathogenic *Escherichia coli* (APEC) are known to carry an impressive arsenal of virulence-associated genes (VAGs), several of which are known to facilitate invasion of epithelial cells and survival in poultry, presumably enabling APEC to disseminate from their initial site of colonization in the respiratory tract to multiple organ sites. This is the first study that describes the genetic composition of drug-resistant APEC in Australia. It indicates that Australian APEC belong to sequence types (STs) that carry a diverse array of VAGs, many of which are highly related to extraintestinal pathogenic *E. coli* recovered from human patients with a variety of associated diseases. VAGs encoding iron acquisition systems, toxins and factors that promote survival in human urine and blood co-residing on ColV-like IncF virulence plasmids were identified in our study. Notably, we identified high carriage rates of IS*26*, an insertion element that is thought to play an important role in the evolution of antimicrobial resistance regions. Further studies are needed to determine the role played by IS*26* in the assembly of complex resistance regions on ColV-like and other APEC plasmids.

4.2.4 Introduction

Escherichia coli are considered the most frequently isolated Gram-negative pathogen affecting human health (80). Over the past 20 years, extraintestinal pathogenic *E. coli* (ExPEC) have risen to prominence. ExPEC colonize the gut asymptomatically, but carry virulence-associated genes (VAGs) that enable them to colonize extraintestinal sites and cause disease. Most ExPEC infections localize to the urinary tract and are known as uropathogenic *E. coli* (UPEC). UPEC can progress from the bladder to cause more serious disease, including pyelonephritis and sepsis (uroseptic *E. coli*).

Another subset of ExPEC, neonatal meningitis-causing *E. coli* (NMEC), can produce severe neurological disease, particularly in infants. ExPEC also cause disease in diverse, agriculturally important, food animal species, particularly poultry (avian pathogenic *E. coli* ; APEC), but also in swine and dairy cattle (93, 115-117). Infections caused by multiple drug resistant (MDR) ExPEC are increasing in frequency and are a major cause for concern (118).

APEC carry large conjugative plasmids containing combinations of iron acquisition genes, including the i*ucABCD*/*iutA* (aerobactin uptake) and *iroBCDEN* (salmochelin uptake) operons, as well as other heavy metal transporters *sitABCD* and *etsABC*, and the serum resistance gene *iss*. The carriage of these VAGs has been linked to the capacity of APEC to cause disease, but their presence is not essential for extraintestinal infection in an avian host (95, 119-123). APEC are genetically heterogenous and carry diverse combinations of VAGs involved in iron acquisition, cytotoxicity, adhesion, invasion and immune evasion. Many of the putative virulence genes found in APEC are also found in human ExPEC. APEC can grow in human urine, resist mammalian complement and invade human epithelial cells (124-127).

Moreover, APEC and human ExPEC share serotypes, sequence types (STs) and pulse-field gel electrophoresis profiles (128-130). Collectively, these and other observations underpin the hypothesis that poultry-associated *E. coli* pose a zoonotic threat (91, 117, 126, 130-132), although the zoonotic potential of APEC has yet to be quantified (74, 133). While animal models seeking to determine the zoonotic potential of APEC have been informative, they have not been definitive. Reverse zoonotic episodes where MDR pathogens carried by humans are transferred to poultry and other food animals also pose a significant biosecurity threat (134).

APEC are found in the intestinal flora of healthy commercial bird species, can cause disease at various anatomical sites, and are a leading cause of mortality and morbidity in poultry of all ages (93, 135). APEC infect the trachea and air sacs, as well as the oviduct, pleura, peritoneum and pericardium, liver, blood, yolk sac, growth plates and joints (136). Many systemic infections caused by APEC are initiated by colonization of the respiratory tract after inhalation of faecally contaminated dust, with subsequent dissemination to more distant sites. However, the factors that precipitate invasion are not well understood (91, 92, 136-138).

The introduction of APEC plasmids into avian commensal *E. coli* has been shown to confer virulence in animal models of ExPEC disease (102). APEC are often resistant to a range of antimicrobial agents, including tetracyclines, chloramphenicol, sulphonamides, aminoglycosides, fluoroquinolones and β-lactams (97) and the corresponding resistance genes are often plasmid-associated. The introduction of MDR plasmids from poultry via

food into the human gut is a potential threat to human health. Whole genome sequencing (WGS) approaches will help to provide insights into the zoonotic potential of APEC and into the role of mobile DNA in pathogen evolution and assembly and the spread of antimicrobial resistance genes (139, 140), although WGS studies of APEC are in their infancy (72), particularly within Australia.

Here, we used WGS to characterize 95 geographically diverse APEC strains that had been determined by PCR to carry a class 1 integrase (*intI1*) gene, a reliable proxy for MDR (31), which here is defined as carriage of three or more genes associated with resistance to different classes of antibiotic. The genome sequences were interrogated for the Clermont phylogroup, e-serotype, multi-locus sequence type (MLST) and VAGs, to seek novel insights into the genetic characteristics of Australian APEC carrying multiple antimicrobial resistance genes.

4.2.5 Methods

Sample origins and associated metadata

The APEC that were investigated had diverse origins. They were obtained between 2007 and 2015 from at least 12 Australian agricultural poultry operations across four states (Victoria and New South Wales, Queensland and Western Australia), although as the geographical origins of some isolates were unclear, the exact number of sources cannot be determined. The APEC originated predominantly from broiler and layer chickens, and to a lesser extent from turkeys and ducks (Table S1 [Additional File 4.1]). Data on any antimicrobial therapy administered to these animals are limited, but in Australia relatively few antimicrobials are approved for commercial poultry and the administration of several active ingredients (e.g. gentamicin, fluoroquinolones and chloramphenicol) is not permitted and therefore extremely unlikely to have been used in the flocks. Samples identified to have originated from the same geographical site at the same time and that shared identical phylogenetic classifications and genotypes were considered duplicate isolates and removed from the analysis.

Isolate collection

Swabs were collected from multiple anatomical sites by a team of experienced veterinarians from the University of Melbourne from deceased or culled birds with signs

of an APEC infection. Anatomical sampling sites varied between birds, but in most cases samples were taken from internal organs. *E. coli* were cultured on sheep blood and MacConkey agar and a routine PCR (78) was used to determine whether the *E. coli* carried the typical repertoire of APEC VAGs. The isolates were stored at -80°C in 20% glycerol or on Protect (Thermo-Fisher) beads.

Determination of intI1 carriage by PCR

Single APEC colonies were picked from LB agar plates and inoculated in 5 ml of LB medium to prepare glycerol stocks and crude DNA templates for PCR (28). Primers HS915/HS916, which span a 371 bp region of *intI1*, were used to identify isolates carrying a class 1 integron, as previously described (141). Isolates that yielded a band of 371 bp amplicon indicative of the *intI1* gene were selected for WGS.

DNA extraction, WGS and assembly

Genomic DNA was extracted using the ISOLATE II Genomic DNA Kit (Bioline) following the manufacturer's instructions for bacterial cells and stored at -20°C. Library preparation was undertaken using Nextera DNA Library Preparation kits generating 150 bp paired end reads from 0.5 ng of template DNA. WGS of strains was performed using an Illumina HiSeq 2500. Sequence read quality was assessed using FastQC version 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) before Illumina raw reads passing quality control were assembled into draft genome sequences using the A5 assembly pipeline version A5-miseq 20150522 (142). Genomes with an average read depth of ≥20, and that also assembled to 600 or fewer scaffolds, were retained for further phylogenetic analysis using Phylosift. Draft genome assemblies were deposited in NCBI; individual accession numbers can be found in Table S1 (Additional file 4.1).

Genotyping and phylogenetic classification

Publicly available databases such as PlasmidFinder, ResFinder, VirulenceFinder (http://www.genomicepidemiology.org/) and ISfinder (143) were used to source reference sequences for genotyping, with additional sequences of interest not present within these databases collected from the NCBI nucleotide database and the Virulence Factor Database (144). Genotyping, including for the purposes of phylogroup (66) and e-serotype classification (145), and MLST (http://mlst.warwick.ac.uk/mlst/) were

performed using the read-mapping tool ARIBA (146) before the processing of such data with a bespoke script accessible on GitHub (https://github.com/maxlcummins/APEC-MGEN-2018).

Single nucleotide polymorphism analyses

Phylogenetic SNP analysis and identification of SNPs in *gyrA* and *parC* conferring fluoroquinolone resistance (*gyrA* : Ser-83-Leu, Asp-87-Asn; *parC* : Ser-80-Ile, Glu-84-Gly) (147) was performed using Snippy version 3.2 (148), with a K12 strain used as a reference (accession no. KU00096.3), and manually curated through use of AliView(149).

Phylogenetic analyses

Maximum-likelihood phylogenetic tree analyses were performed under a generalized time-reversible model using the PhyloSift pipeline version 1.0.1 (150) and FastTree version 2.1.8 (151), altered to resolve short branches as previously described (152), and visualized in iTOL (153). Phylogenetic SNP trees were generated using Snippy, with the resulting full core genome alignment filtered for recombination with Gubbins (https://sanger-pathogens.github.io/gubbins/). SNP-sites v2.4.0 (https://github.com/sangerpathogens/snp-sites) was then used to create an alignment consisting of 1539 variable sites before tree generation with FastTree, also using a generalized time-reversible model. This tree was visualized using the R package ggtree (154). SNPs were counted using snpiphy (https://github.com/bogemad/snpiphy). Additional details are available at https://github.com/maxlcummins/APEC-MGEN-2018.

Inference of ColV-like virulence plasmid carriage

Short reads from each sample were mapped to the reference plasmid pCERC4 (accession no. KU578032) using the Burrows–Wheeler Aligner (BWA) 0.7.17 (155) and converted to a BAM file format using SAMtools 0.1.18 (156). Through use of a bespoke Python script each BAM file was then used to produce a histogram of read-depth as a function of reference coordinate, clustered based on their euclidean distances, and used to generate a heatmap. A schematic of pCERC4 was then generated using SnapGene (https://www.snapgene.com/) and overlaid above the heatmap to facilitate visualization of the relative genetic loci where sample reads were mapped. The scripts

and commands used are accessible on GitHub (https://github.com/maxlcummins/APEC-MGEN-2018).

4.2.6 Results and Discussion

ST117 and ST350 are predominant lineages

Out of 256 APEC isolates, 123 were predicted by PCR to carry a class 1 integron and were sequenced. Twenty-six isolates failed to be sequenced and/or assembled to a quality that met our aforementioned assembly criteria, while an additional two isolates were considered duplicates and removed from further analysis. Assembly statistics of samples analysed are shown in Table S1 (Additional file 4.1). Based on comparisons of 37 core protein sequences from PhyloSift, these remaining 95 APEC clustered into five clades and were heterogeneous in nature (Figure 4.1). Genomes sharing the same ST and e-serotype clustered, as expected. Analysis revealed that phylogroup D (58/95) and, to a lesser extent, B2 (21/95) were dominant in Australian *intI1*-positive APEC populations, cumulatively representing 82% (79/95) of the collection. The distribution of APEC between phylogroups appears to vary in different countries. While APEC within phylogroups B2 and D (72, 157) usually predominate, some APEC populations can also contain high proportions of members of phylogroup A (88). The increased frequency of typically commensal phylogroups A and B1 among ExPEC and APEC (72) suggests that pathogenic *E. coli* continue to evolve rapidly via mechanisms underpinned by lateral gene transfer.

Thirty STs were identified among the 95 APEC. There were several dominant lineages, including ST117, clonal complex 350 (CC350) (ST350, ST57), ST95, ST429 and ST973, which cumulatively represented 64 of the 95 APEC. Notably, ST117 was the most frequent ST (22/95) in the collection. ST117 is a major cause of extraintestinal infections in poultry, is considered an emerging human pathogen globally (73, 99), and is noted for its carriage of genes encoding extended-spectrum β-lactamases (158). APEC-associated ST117 isolates are serotypically diverse, but most have H4 flagella (72, 159, 160). Among our 23 CC117 isolates (ST117, 22 isolates; ST4045, one isolate), seven serotypes were identified, but serotypes O78:H4 and O111:H4 predominated. These serotypes have also been identified in studies of APEC in Brazil and several Nordic countries (99, 161-163), suggesting that they may represent globally disseminated subclones.

Although the genotypic profiles of the 23 CC117 isolates were variable, all were rich in VAG content, particularly those in serogroups O78 and O111 (Figure 4.2). All CC117 isolates carried *etsA*, *iucD* /*iutA* and *sitA*, while *ireA* (22/23, 96 %), *iroN* (21/23, 91 %) and *fyuA* (20/23, 87 %) were also common. The protectin-associated genes *iss* and *ompT* were also ubiquitous in these isolates, and 91% (21/23) also carried *traT*. Genes thought to be involved in host-cell adhesion were also common (*irp2* and *tia*, 9/23, 83 %; *papGII*, 17/23, 74 %). Plasmid replicon types within this CC were quite diverse, with *repA* genes associated with nine different incompatibility groups represented. The most common were FIB (23/23, 100 %), FII (22/23, 96 %) and HI2 (4/23, 17 %). IncN (2/23, 9 %), IncI1 (2/23, 9 %), IncI2 (1/23, 4 %) and IncY (1/23, 4 %) were also detected.

The mean and median SNP counts across all CC117 isolates, relative to 2009–3133, were 277 and 348, respectively, with 57% (13/23) of the isolates having 50 or fewer SNPs when compared to one or more CC117 isolates within the collection. While some of the CC117 isolates of the same serotype exhibited low SNP counts across their core genome, there were also examples of serotypically homogenous isolates that had high SNP counts (Figure 4.3; Table S2 [Additional file 4.2]). For example, isolates AVC77 and AVC222 were both serotype O111:H4 but differed from each other by only 23 SNPs, even though AVC77 was isolated from a bird from Western Australia in 2008 and AVC222 was isolated from a bird from Victoria in 2012; these are states separated by more than 1000 km. In contrast, two other isolates with the same serotype, O78:H4 (AVC96 and AVC29), were collected from these same states just one year apart and yet differed by 497 SNPs.

Apart from the 23 isolates of CC117, other prevalent lineages were ST57 (9/95; 9.5 %) and ST350 (10/95; 10.5 %), which belong to the same CC (CC350). The 19 members of this CC all carried *ompT*, *iucD* /*iutA*, *iss*, *iroN* and *hylF*. Other VAGs identified within this population included the pap operon, found in 37% (7/19) of samples, and the adhesin *tsh*, which was found in 79% (14/19) of isolates. APEC belonging to CC350 are frequently isolated from poultry in Australia and in other countries (157, 164, 165), are associated with extraintestinal infections including urinary tract infections (UTIs) and sepsis (165, 166), and may constitute a potential poultry-associated zoonotic agent. Two ST57 isolates from the collection under investigation were each found to carry a total of nine

AMR genes, which, in combination with their extensive VAG profiles, highlighted them as a potential emerging pathogen.

Our APEC collection included 8/95 isolates belonging to ST95 and there were three different serotype profiles among them. These isolates had the highest level of carriage of extraintestinal VAGs; all carried *iroN* , *iss* , *kpsMT*(II), *iucD* , *ompT*, *papGII* , *neuC* and *usp*, while the gene encoding the vacuolated autotransporter toxin Vat and *gimB*, a marker of the genetic island linked to neonatal meningitis *E. coli* (91) and a capacity to invade cells (167), were carried by 6/8 of these isolates. ST95 is well documented in the literature as an APEC lineage that also frequently appears as a causative agent of UTIs and blood sepsis in humans (157, 168, 169).

The antimicrobial resistance (AMR) gene profiles of the eight ST95 isolates were variable, with one isolate carrying six AMR genes and the remaining isolates carrying only one or two. The ST95 lineage is unusual in that isolates are reported to have a lower level of acquired resistance than other pandemic lineages (169). Fully assembled genomes of MDR ST95 contain resistance genes linked to large IncFIB/IncFII plasmids (170). Plasmids with IncFII and IncFIB *repA* genes predominated in our collection and all eight ST95 isolates carried these markers. These data suggest that ST95 APEC of Australian origin, despite being variable in their AMR profiles, carry a significant reservoir of VAGs associated with human ExPEC infections, including UTIs, septicaemia and neonatal meningitis, and may have zoonotic potential.

intI1-positive Australian APEC do not carry antimicrobial resistance genes of major clinical significance

We purposely targeted APEC that carried a class 1 integrase for WGS to maximize the likelihood of characterizing MDR strains. The most frequently identified AMR genes in the collection were *sul1* (59/95, 62 %), *tet(A)* (51/95, 54 %), $bla_{\text{TEM-1A/B/C}}$ (48/95, 51 %), *dfrA5* (48/95, 51 %), *strAB* (36/95, 38 %), *aadA1* (32/95, 34 %), *sul2* (31/95, 33 %), *dfrA1* (25/ 95, 26 %), *tet(C)* (18/95, 19 %), *tet(B)* (12/95, 13 %) and *aadA2* (4/95, 4 %) (Figure 4.2).

The AMR genotypic profiles of these APEC isolates reflect the antimicrobial stewardship practices used widely in Australian poultry production systems. The most common

phenotypic antimicrobial resistances reported in Australian APEC are to tetracycline, trimethoprim/sulfamethoxazole, streptomycin and ampicillin, at prevalences of 75, 38, 22 and 9%, respectively (171). Our findings are largely consistent with these phenotypic resistance data, although we would have expected higher rates of resistance to ampicillin in the literature given the carriage rates of bla_{TEM-1} in the APEC sequenced here.

Only one isolate, AVC111-ST354-ONT:H34, was found to have SNPs in *gyrA*/*parC* associated with fluoroquinolone resistance (FQR). This sample is of an ST which was reported in a study on FQR *E. coli* from canine faeces and cases of human ExPEC infection as a dominant lineage between both such sources (172), and therefore these *gyrA*/*parC* mutations are probably clonal. Additionally, analysis of APEC and Avian faecal *E. coli* (AFEC) in Australia identified strains of ST354 with FQR (173). Otherwise, the APEC in our collection did not carry genes conferring resistance to antimicrobials important in the treatment of human disease, including cephalosporins, fluoroquinolones, carbapenems and colistin, an observation in stark contrast to those made on APEC isolated in many other countries. Studies in China and Egypt have reported that 75 and 23% of APEC carry bla_{CTX-M} genes and bla_{SHV} genes (174, 175), respectively, while several studies on APEC from South Africa, China, Egypt and Vietnam have also identified APEC isolates carrying *mcr-1* (176-179).

We also failed to find any evidence of the carriage of genes encoding resistance to cephalosporins, fluoroquinolones, carbapenems or colistin among the genome sequences of porcine commensal *E. coli* that carry class 1 integrons (28). This highlights the benefits of enforcing legislation to control use of critically important antimicrobials in food animals, as colistin, gentamicin, fluoroquinolones and amphenicol antimicrobials are not registered for use in food production animals in Australia (although restricted use of cephalosporins is allowed (180)). On a cautionary note, we detected IS*26* at a high prevalence in our APEC collection. IS*26* is an insertion element that forms composite transposons carrying a wide variety of antimicrobial resistance genes (15, 181), promotes cointegrated plasmid formation (182) and enhances plasmid fitness (183). IS*26* can also recognize existing copies of IS*26* (184, 185) and promote formation of complex resistance gene regions (186, 187). Therefore, long read sequencing would be

useful in the investigation of the genetic context of the AMR genes detected, and other regions that abut insertion sequences such as IS*26*.

Carriage of virulence-associated genes in APEC

Carriage of VAGs among Australian *intI1*-positive APEC isolates is shown in Figure 4.2. Genes encoding iron capture systems were frequently represented. Specifically, *iutA* and *iucD* (aerobactin operon), *iroN* (salmochelin operon) and *sitA* (sit operon) were often detected (84/95, 88 %; 84/95, 88 %; 84/95, 88 %; and 88/95, 93 %, respectively), while carriage of *ireA*, *irp2* and *fyuA* (Yersiniabactin operon) was less common (56/95, 59 %; 41/95, 43 %; 42/95, 44 %, respectively). The prevalence of these VAGs in Australian APEC is similar to that seen in APEC from other countries (78, 188-190). Iron is tightly held in mammalian tissues and is a major factor limiting the growth of pathogens. APEC have evolved complex strategies, including the expression of specialized siderophores and iron chaperones, to recover iron from their host (135).

VAGs mediating protection against complement resistance are thought to be essential for the ability of APEC to disseminate to extrapulmonary sites. Almost all 95 APEC carried the increased serum survival gene *iss* (91/95, 96 %), *ompT* (91/95, 96 %) and *traT* (90/95, 95 %), genes that have been epidemiologically associated with or determined experimentally to confer serum resistance in ExPEC (87, 191-193). A recent study on 50 Australian APEC reported identical carriage rates of *iss* and *ompT* (173). The importance of the *iss* gene as a marker of APEC is reinforced by its inclusion in a diagnostic pentaplex PCR (78). Moreover, the Iss protein has been trialled as a heterotypically protective antigen in an experimental APEC vaccine (194). More than a third (36/95) of the APEC isolates carried a variant of *kpsM*, the product of which is known to mediate complement resistance, a key characteristic of APEC. Group II variants of *kpsMT* are frequently detected among human ExPEC but are less commonly detected in APEC globally. This locus was detected only within APEC in phylogroups B2 and D, which are historically associated with extraintestinal disease, with the notable exception of one ST93 isolate within phylogroup B1.

Many APEC-associated adhesins have been described, but their presence is not exclusive to APEC, so adhesin genes are poor diagnostic markers for APEC (95, 130). All APEC in this study carried the fimbrial adhesin gene *fimH*, while 43% (41/95) carried the putative

adhesin *tsh* and 34% (33/ 95) of the APEC isolates in this study carried *papGII* (pyelonephritis-associated pilus tip adhesin gene). Pap is thought to play a role in systemic extraintestinal infections of poultry and colonization of the kidneys in humans and the reproductive tract of dogs (195-199).

Australian APEC frequently carry IncFIB and IncFII

Carriage of at least one plasmid *repA* gene was common in the Australian APEC studied here (Figure 4.2). The most common *repA* genes belonged to Inc-types FII and FIB, which were each present in 98% (93/95) of our APEC collection. Inc FII and FIB are commonly found in APEC globally (176) and are associated with large conjugative virulence plasmids that are a feature of the APEC phenotype (177). IncB/O/K/Z and IncI1 incompatibility marker genes were also detected frequently (25/95, 26 %; 24/95, 25%; respectively). An investigation by Johnson et al. in 2007 (200) found IncB/O/K/Z replicons at a similar prevalence (24 %) among a collection of 422 APEC and detected IncI1 at a slightly higher prevalence of 41%. IncHI2, IncI2 and IncN *repA* genes were detected in 8% (8/95), 7% (7/95) and 4% (4/95) of our Australian APEC isolates, respectively.

Prevalence of ColV-like virulence plasmids

A preliminary analysis of the frequency of virulence plasmid-associated VAGs in our collection suggested that ColV-like plasmids were a feature of these Australian APEC. To examine this further we used a recently published ColV plasmid sequence (KU578032) as a reference to map Illumina short-reads derived from our 95 APEC genome sequences. A bespoke python script was used to construct a visualization of the coverage of mapped APEC reads (Figure 4.4). The utility of this approach is demonstrated by the observation that all 11 APEC isolates that did not carry *iutA* and *iucD*, as determined by ARIBA (e.g. isolates AVC51 and AVC207), lacked reads mapping to the corresponding region of the reference plasmid pCERC4, as shown in Figure 4.4. Our analyses also showed that genomes belonging to the same ST and e-serotype shared high similarity in their read mapping profiles and therefore commonly clustered together, suggesting they may carry closely related plasmids.

Along with the high carriage of virulence plasmid-associated VAGs and IncFIB/IncFII repA genes, our data suggest that ColV/ColBM-like plasmids are common in *intI1*-positive Australian APEC, an observation that mirrors studies elsewhere (200). However, while Figure 4.4 strongly suggests that the VAGs are within a plasmid, the analysis allows mapped reads to be recruited from any part of the genome and does not confirm the co-localization of the VAGs on a ColV-like plasmid backbone.

Notably, reads from samples AVC103, AVC171 and AVC173 all mapped extensively across the reference sequence pCERC4, which is sourced from a human commensal ST95 strain (201). All such samples were also identified as ST95; this ST is well documented in associations with poultry meats, poultry disease and human extraintestinal infections (169). Therefore, it is possible that this strain and/or plasmid may be closely related to those of the ST95 APEC samples under study. Long read sequencing of these samples and other Australian APEC would assist in the determination of VAG and AMR gene context and allow for comparative genomic investigations that may infer the movement of microbial populations and/or their plasmid content between different environmental contexts.

It is important to note two limitations of the study: sampling was inconsistent by state, and our collection is biased through selection based on carriage of *intI1*. Despite these limitations our study suggests: (i) that while *intI1*-positive Australian APEC are phylogenetically and serotypically diverse, particular lineages, such as CC117 and CC350, appear to constitute the primary health burden in the poultry sector; and (ii) these APEC carry large virulence plasmids which may also harbour genetic elements conferring resistance to antibiotics used to treat UTIs, such as trimethoprim and sulfamethoxazole, as well as genes encoding resistance to a wide array of first-generation antibiotics. Further work is required to investigate the genetic context of the AMR genes described here, and regions that abut insertion sequences such as IS*26*, because many AMR genes are mobilized by IS*26* (13, 202). While many of the APEC isolates under investigation were determined to be genotypically MDR, none carried genes conferring resistances to critically important antibiotics such as colistin, extended spectrum β-lactams or fluoroquinolones, except one sample that carried SNPs linked with resistance to fluoroquinolones. A subset of APEC, such as ST117 and ST95, are phylogenetically and

genotypically similar to *E. coli* that cause human extraintestinal infections, highlighting a potential zoonotic risk. Efforts are needed to ensure poultry are restricted in their capacity to be a reservoir of pathogenic *E. coli*, particularly those that may pose a zoonotic risk and carry broad-host conjugative plasmids containing VAGs and AMR genes.

4.2.7 Figures

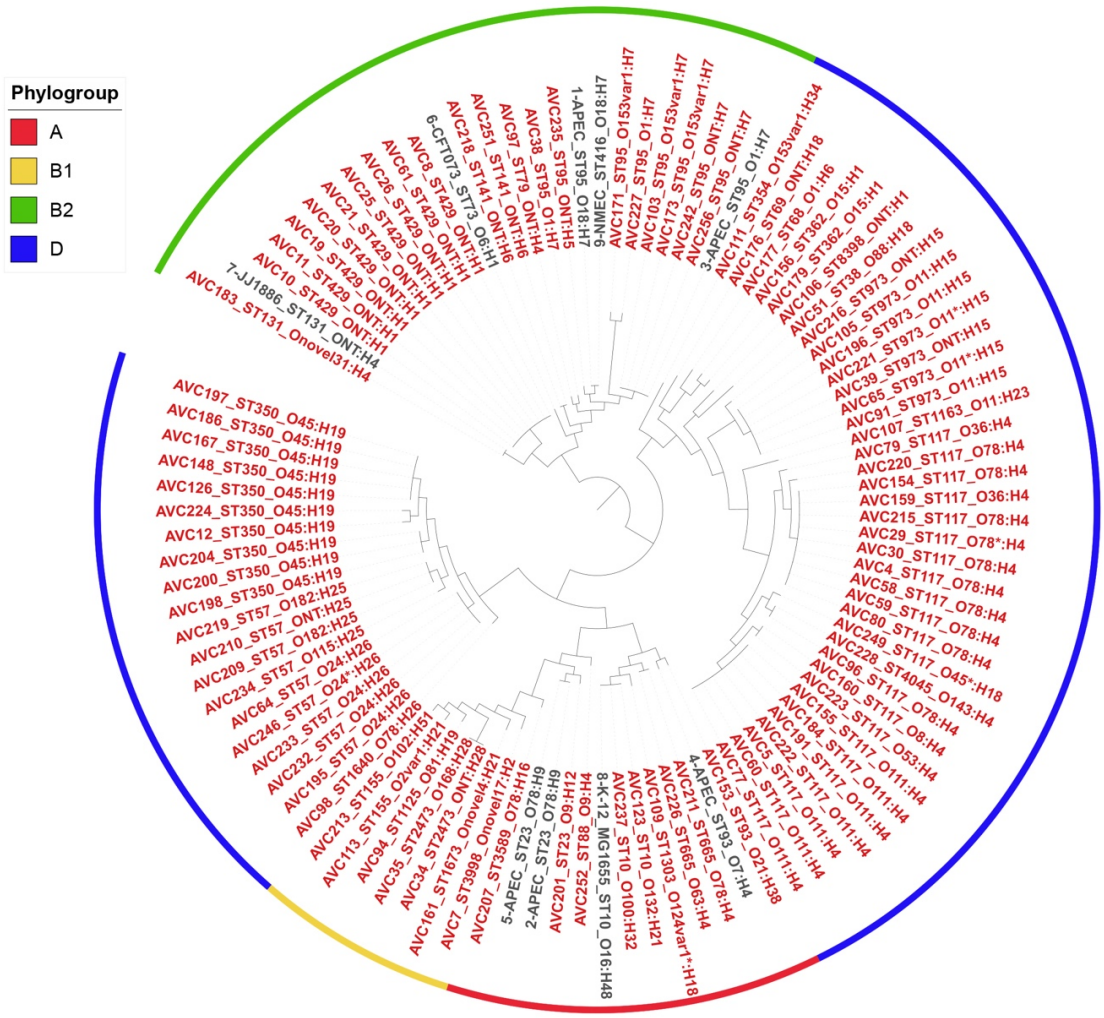

Figure 4.1 - Phylogenetic relatedness of the APEC isolates and nine reference strains (accession numbers: 1, CP006830; 2, CP007442; 3, NC_008563; 4, CP013048; 5, CP004009; 6, AE014075; 7, CP006784; 8, U00096; 9, CP007275), as determined by PhyloSift in combination with FastTree. The tree is midpoint rooted. Red tip labels indicate isolates from the present study, while those shown in black are the reference strains. The STs and e-serotypes of isolates have been appended to their names (e.g. A5_ST117_O111:H4). The coloured bands encircling sections of the tree indicate the phylogroups into which isolates were categorized. Asterisks in sample names indicate low read depth at one or more loci for a given MLST/O-type/H-type (see https://www.github.com/maxlcummins/ARIBAlord for more information).

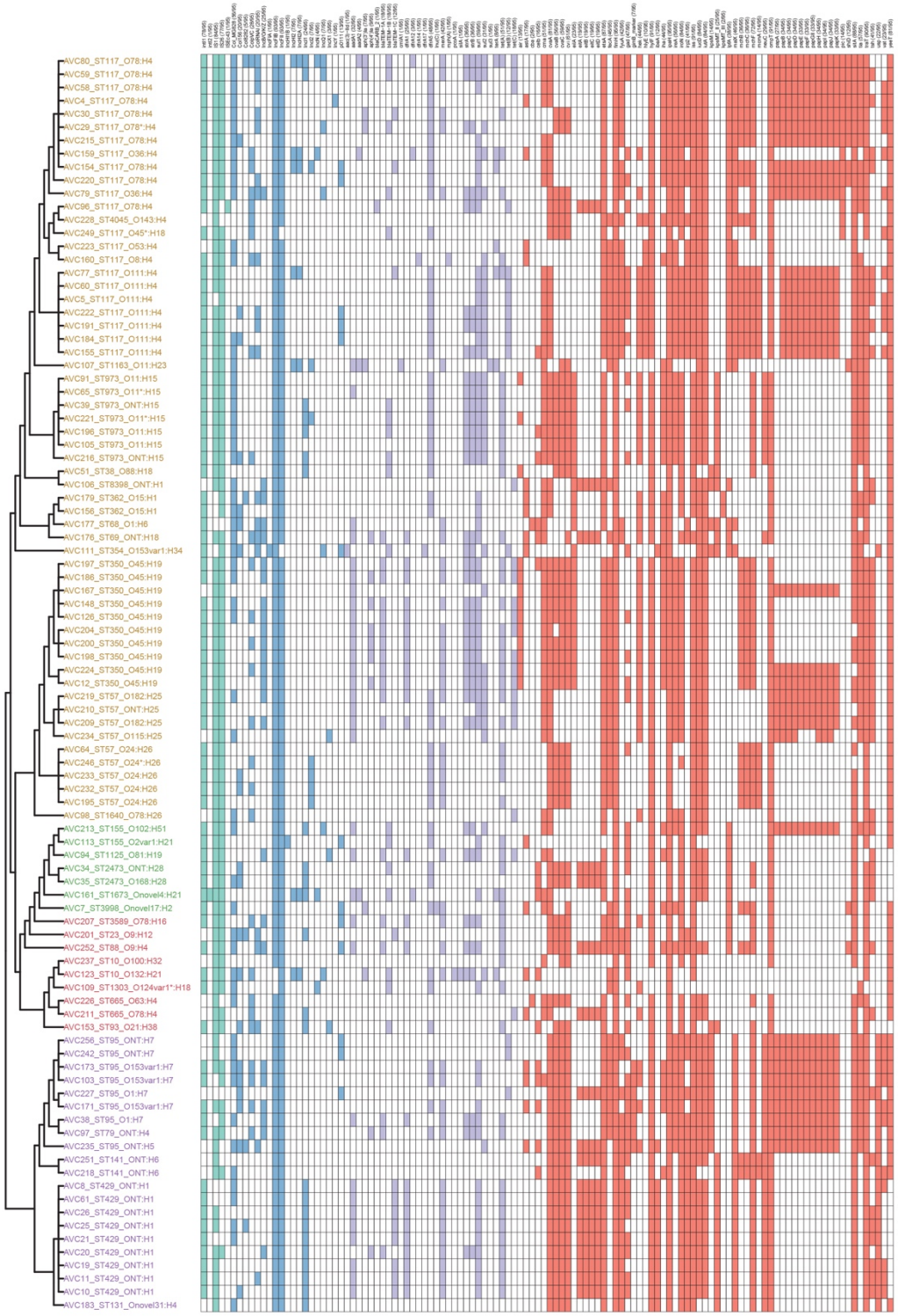

Figure 4.2 - Genotypic profiles of APEC isolates, clustered on the basis of the PhyloSift tree in Figure 4.1, with the tip labels indicating the ST and e-serotype (in text) and phylogroup (indicated by text colour, where A is red, B1 is green, B2 is purple and D is mustard). Carriage of mobile genetic element-associated genes (teal), AMR genes (purple), VAGs (red) and plasmid *repA* genes (blue) are shown adjacent to the tree in a hit-table, with a white square indicating the absence of a specific gene.

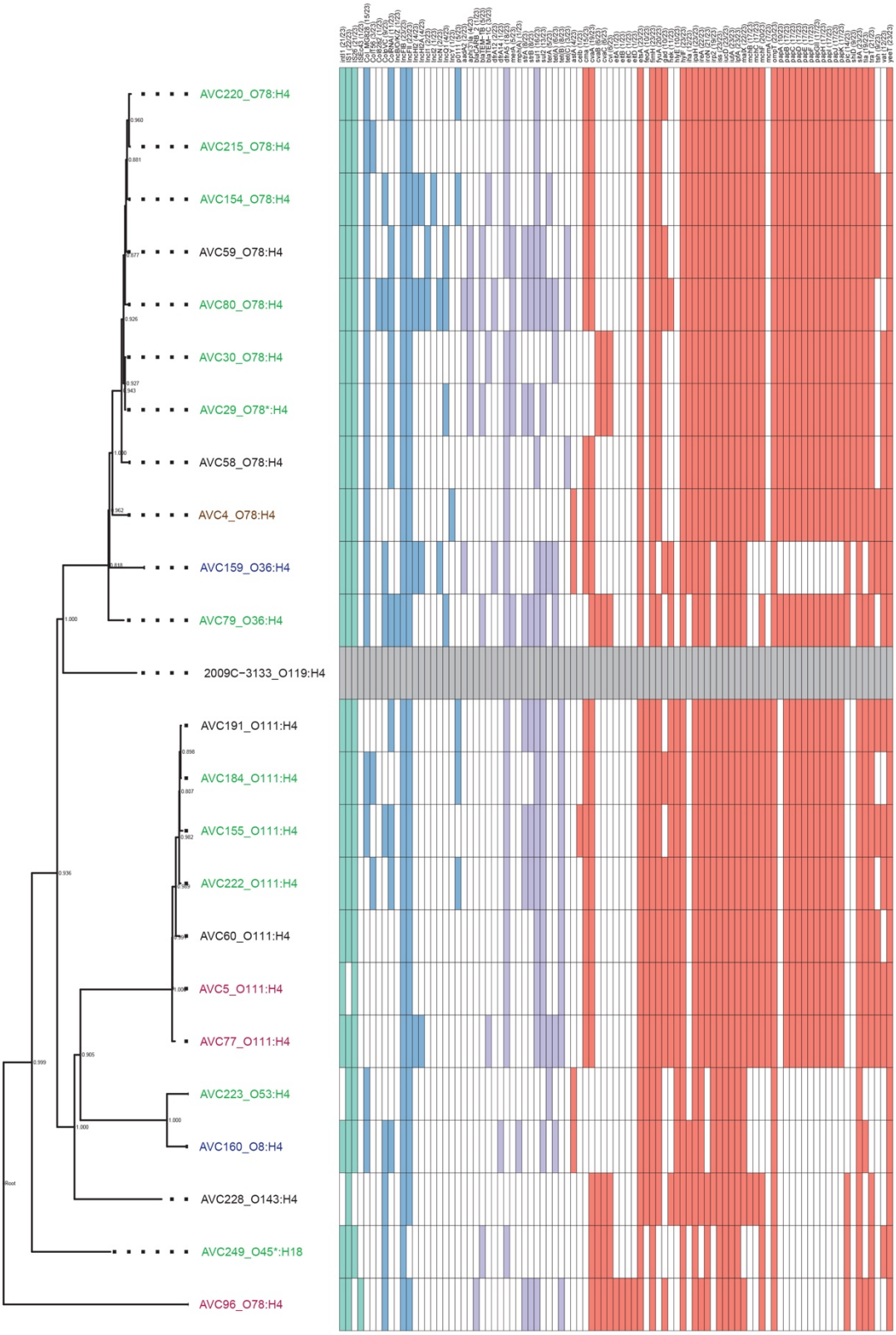

Figure 4.3 - High-resolution phylogenetic comparison of ST117 APEC isolates, as determined by Snippy in combination with Gubbins, SNPsites and FastTree. The tree is midpoint-rooted. The ST and e-serotype are shown on the tip labels, while the colour of the labels indicates the state of origin (brown, QLD; green, VIC; red, WA; blue, NSW; black, unknown). A gene hit map is also shown; mobile genetic element-associated genes (teal), AMR genes (purple), VAGs (red) and plasmid *repA* genes (blue) are shown adjacent to the tree in a hit-table, with a white square indicating the absence of a specific gene.

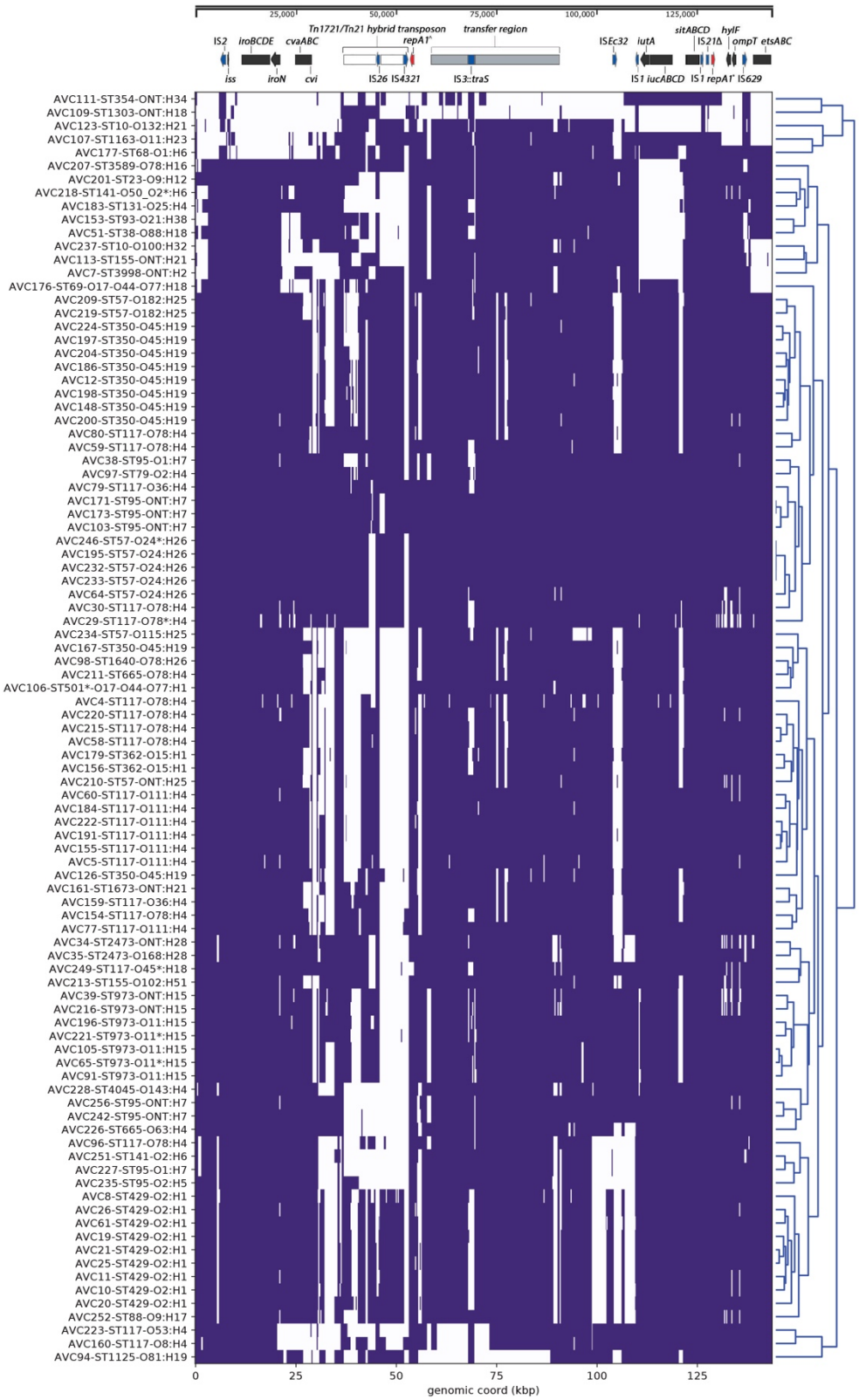

Figure 4.4 - Mapping of short-reads indicating the presence of ColV-like virulence plasmids. Purple colour indicates a median depth of 10 or greater at a given 250 bp bin, whereas white space indicates the inverse. Clustering of rows on this heatmap is based on similarity between the coverage profiles of the isolates, while a schematic of pCERC4 is shown above the heatmap to provide an overview of the genetic elements that were present or absent based on this analysis. Key: *repA1*^, IncFII *repA1* gene; *repA1**, IncFIB *repA1* gene.

Chapter 5: *Salmonella* Genomic Island 1B Variant Found in a Sequence Type 117 Avian Pathogenic *Escherichia coli* Isolate

Authors

Max Laurence Cummins[a], Piklu Roy Chowdhury[a], Marc Serge Marenda[c], Glenn Francis Browning[b], Steven Philip Djordjevic[a*]

Author affiliations

[a] The ithree institute, University of Technology Sydney, Ultimo, New South Wales, Australia

[b] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[c] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Werribee, Victoria, Australia

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Published in mSphere, 22 May 2019. https://doi.org/10.1128/mSphere.00169-19.

5.1 Declaration

M.L.C. was responsible for project administration, conceptualisation, and investigation, writing and execution of software, formal curation, analysis, validation and visualisation of data and writing the original manuscript draft and editing.

P.R.C: responsible for formal analysis and reviewing and editing draft works.

M.S.M, G.F.B: responsible for sample collection and metadata collection.

S.P.D. supervision, resource provision, formal analysis, reviewing and editing draft versions. All authors have read and agreed to the published version of the manuscript.

5.2 *Salmonella* Genomic Island 1B Variant Found in a Sequence Type 117 Avian Pathogenic *Escherichia coli* Isolate

Please note that this chapter is relatively short as it was published as an 'observation' format and therefore has a combined introduction, results and discussion. It addresses Aims 1 and 2 of the thesis. Please note that figure numbers have been changed to

maintain consistency with the thesis as a whole. Methodology is provided as an appendix (File S5.1) as it was provided as Supplementary material for publication. Figure 6.1 is better viewed via the additional files submitted with this thesis.

5.2.1 Abstract

Salmonella genomic island 1 (SGI1) is an integrative genetic island first described in *Salmonella enterica* serovars Typhimurium DT104 and Agona in 2000. Variants of it have since been described in multiple serovars of *S. enterica*, as well as in *Proteus mirabilis*, *Acinetobacter baumannii*, *Morganella morganii*, and several other genera. The island typically confers resistance to older, first-generation antimicrobials; however, some variants carry $bla_{\text{NDM-1}}$, $bla_{\text{VEB-6}}$, and $bla_{\text{CTX-M-15}}$ genes that encode resistance to frontline, clinically important antibiotics, including third-generation cephalosporins. Genome sequencing studies of avian pathogenic *Escherichia coli* (APEC) identified a sequence type 117 (ST117) isolate (AVC96) with genetic features found in SGI1. The complete genome sequence of AVC96 was assembled from a combination of Illumina and single-molecule real-time (SMRT) sequence data. Analysis of the AVC96 chromosome identified a variant of SGI1-B located 18 bp from the 3` end of *trmE*, also known as the *attB* site, a known hot spot for the integration of genomic islands. This is the first report of SGI1 in wild-type *E. coli*. The variant, here named SGI1-B-Ec1, was otherwise unremarkable, apart from the identification of ISEc43 in open reading frame (ORF) S023.

5.2.2 Importance

SGI1 and variants of it carry a variety of antimicrobial resistance genes, including those conferring resistance to extended-spectrum β-lactams and carbapenems, and have been found in diverse *S. enterica* serovars, *Acinetobacter baumannii*, and other members of the Enterobacteriaceae. SGI1 integrates into Gram-negative pathogenic bacteria by targeting a conserved site 18 bp from the 3` end of *trmE*. For the first time, we describe a novel variant of SGI1 in an avian pathogenic *Escherichia coli* isolate. The presence of SGI1 in *E. coli* is significant because it represents yet another lateral gene transfer mechanism to enhancing the capacity of *E. coli* to acquire and propagate antimicrobial resistance and putative virulence genes. This finding underscores the importance of whole-genome sequencing (WGS) to microbial genomic epidemiology,

particularly within a One Health context. Further studies are needed to determine how widespread SGI1 and variants of it may be in Australia.

5.2.3 Data availability

Long-read whole-genome sequence data and short-read whole-genome sequence data are available in the SRA under accession no. SRR8671292 and SRR7469869, respectively, while the nucleotide sequence of SGI1-B-Ec1 is available on the NCBI nucleotide database under accession no. MK599281.

5.2.4 Introduction, Results and Discussion

Salmonella genomic island 1 (SGI1) is a site-specific, integrative genetic element that uses a tyrosine recombinase encoded by int_{SGI1} to target the terminal 18 nucleotides (*attB*) of the *trmE* (formerly *thdF*) gene, which encodes a highly conserved GTPase (34). A toxin-antitoxin system (*sigAT*) encoded within SGI1 plays a critical role in its stable maintenance in the host chromosome (46), and while the island can excise as a circularized form via a process that requires int_{SGI1}(34), the frequency at which this occurs in the wild is thought to be very low and is not well understood (203). The transcriptional regulator complex AcaCD encoded by genes on IncA/C plasmids is sufficient to trigger excision and mobilization of SGI1 (34, 204), yet IncA/C plasmids are not known to coexist in the same host as SGI1, suggesting that an active exclusion mechanism limits opportunities for transposition. These observations in part explain why SGI1 is stably maintained in the chromosome, the difficulties encountered in assaying for circular forms of SGI1 (low abundance), and the apparent low transposition frequency of the island (34, 205). SGI1 comprises a backbone of 27.4 kb and a complex class 1 integron (In*104*) of 15 kb that resides in *resG* (open reading frame [ORF] S027). In*104* is flanked by a 5-bp duplication consistent with transposition into *resG*. Variations in the size of In*104* arise depending on the resistance gene cargo it carries, homologous recombination events between shared sequences within the integron, the presence of other mobile elements, and the action of IS elements (36), particularly IS*26*.

The introduction of IS*26* in SGI1 creates further opportunities for the acquisition of diverse antibiotic resistance genes and the rapid evolution of these elements. Notable in this regard is SGI1-L2, which carries an IS*26*-flanked composite transposon containing

multiple antibiotic resistance genes in S024 (206). IS elements such as ISVch4 (IS*1359*) are associated with deletions in the SGI1 backbone, and these events contribute to the ongoing evolution of the element. SGI1 and variants of it may be able to integrate into a wide variety of Gram-negative bacteria because the sequence of the terminal 18 nucleotides of *trmE* (*attB*) is well conserved (40). Experiments performed in vitro have demonstrated that SGI1 is able to integrate into *Klebsiella pneumoniae* and *Escherichia coli* (207), but evidence of the presence of the island in these species in natural environments has been lacking. Since the identification of SGI1 in *Salmonella enterica* serovar Typhimurium DT104 almost 20 years ago, homologous recombination events, as well as insertion sequence-mediated indels, have led to the emergence of more than 30 SGI1 variants, some of which carry antimicrobial resistance genes that are of major clinical significance (208). SGI1 and variants of it have been detected in diverse serovars of *S. enterica* and other Gram-negative pathogens (36, 37, 209-211). For example, *Proteus* genomic island 1 (PGI1), identified in *Proteus mirabilis*, carries extended-spectrum β-lactamase and/or metallo-β-lactamases (211, 212), and SGI variants have been reported in *Morganella morganii* subsp. *morganii* (208), *Acinetobacter baumannii* (38), *Enterobacter hormaechei* subsp. *oharae* (213), and *Providencia stuartii* (214).

While performing an in silico analysis of whole-genome sequencing (WGS) data from 95 Australian avian pathogenic *E. coli* (APEC) isolates (215), one isolate (AVC96) from a diseased 26-week-old broiler chicken was found to carry genetic signatures typically found in SGI1 (GenBank accession no. AF261825). Details of the materials and methods used for analysis of the isolate are given in Text S1 in the supplemental material (Appendix, Supplementary File 5.1). Sequence analysis identified AVC96 as an APEC isolate with sequence type 117 (ST117), a lineage associated with extraintestinal infections in humans and poultry (133). A hybrid assembly using the program Unicycler, which combined Illumina short reads and single-molecule real-time (SMRT) sequences derived from a Pacific Biosystems RSII sequencer, resolved the structure of the SGI1 variant in isolate AVC96 and placed it a single 4,886,273-bp chromosomal contig. The SGI1 variant was inserted in the terminal 18 bp of *trmE*. The variant of SGI1 was here named SGI1-B-Ec1. Comparative analysis with published SGI1 reference sequences revealed that the structure of SGI1-B-Ec1 in isolate AVC96 is related to SGI-1B (accession

no. KU987430), as seen in Figure 5.1. A homologous recombination event between the copies of *intI1* resulted in the loss of the intervening DNA, a feature of this variant. SGI1-B-Ec1 differs from SGI1-B and other SGI1 variants via the insertion of ISEc43 in *S023*. ISEc43 is flanked by an 8-bp direct repeat, suggesting its integration is a recent event. The location of ISEc43 in S023 has not been previously described, and it may serve as a unique epidemiological marker for tracking isolates that carry SGI1-B-Ec1 in Australia. SGI1-BEc1 also carries a unique single nucleotide polymorphism within *qacEΔ1* (228 bp). In *E. coli*, *trmE* sits proximal to *tnaC*, which encodes a tryptophanase. In the case of AVC96, SGI1-B-Ec1 sits between these ORFs. An analysis of 455,632 bacterial whole-genome sequence data sets in the short-read archive (216) indicated that none of the approximately 38,000 *E. coli* genomes available therein carry an SGI1 variant at this locus. BLASTn analysis of the publicly available nucleotide database yielded one entry (GenBank accession no. KU842063.1) that spanned from base 31 of S044 to base 153 of *tnaA*. This sequence was the derived from an *in vitro* experiment that sought to determine the ability of SGI1 to integrate into *E. coli* (207). Therefore, our findings support the contention that AVC96 is the first description of the occurrence of a variant of SGI1 in wild-type *E. coli*. It is notable that variants of SGI1 carrying $bla_{\text{NDM-1}}$ (44), $bla_{\text{VEB-6}}$ and *qnrA1*(43), and $bla_{\text{CTX-M-15}}$ (42) have been identified in multiple drug-resistant *Proteus mirabilis* and *Salmonella enterica* isolates. This discovery should prompt investigations on the prevalence of SGI1-B-Ec1 in Australia and how it might evolve to capture a broader selection of antimicrobial resistance genes.

5.2.5 Acknowledgements

These studies were supported in part by grants from Australian Poultry and Poultry Cooperative Research Centres. We also thank Mohammad Hamidian for contributions in the development of Figure 5.1. This project was partly funded by the Australian Centre for Genomic Epidemiological Microbiology (AusGEM), a collaborative partnership between the NSW Department of Primary Industries and the University of Technology Sydney. M.C. is a recipient of an Australian Government Research Training Program Scholarship. The authors declare they have no conflicts of interest.

5.2.6 Figures

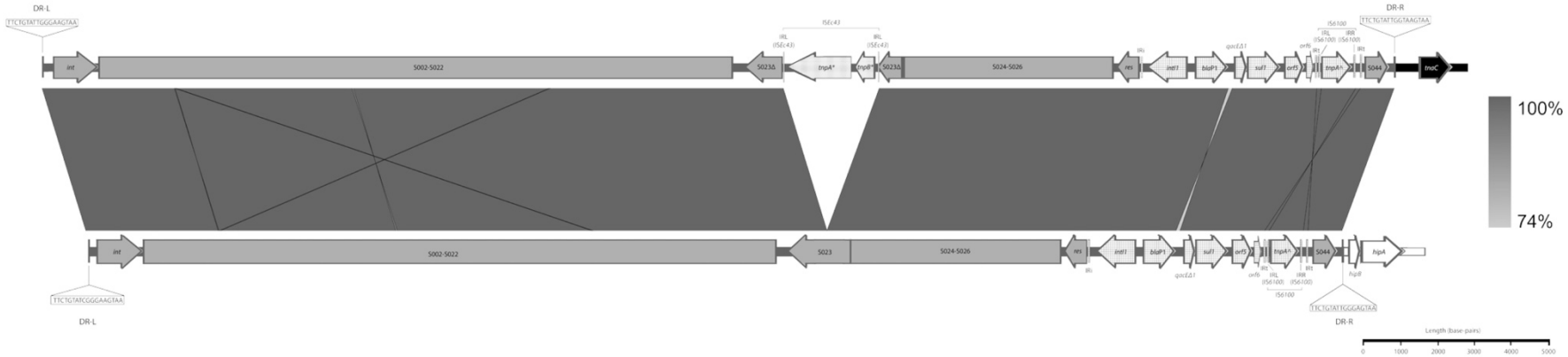

Figure 5.1 - Schematic showing the structural homology between SGI1-B-Ec1 (top) and SGI1-B (bottom). Left and right direct repeats (DR-L and DR-R, respectively) are shown flanking either element. Integron-associated elements are shown with a crosshatched pattern, while other elements of the SGI-1 backbone are shown in dark gray. Genetic elements downstream of SGI1-B-Ec1 are shown in black, while those downstream of SGI1-B are shown in white. ORFs and inverted repeats of IS element ISEc43, unique to SGI1-B, are shown with a dotted pattern near the center of the element. "*tnpA**" and "*tnpB**" are ISEc43 associated, and "*tnpA*^" is IS*6100* associated. Note that genomic coordinates are not to scale and are only approximate. See the GenBank entry for SGI1-B-Ec1 (accession no. MK599281) for precise feature coordinates.

Chapter 6: *Salmonella* Genomic Island 1 is Broadly Disseminated within Gammaproteobacteriaceae

Authors

Max L. Cummins[1,2], Mohammad Hamidian[1] and Steven P. Djordjevic[1,2*]

Author affiliations

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

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

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Published in Microorganisms, 20 January 2019.
https://doi.org/10.3390/microorganisms8020161

6.1 Declaration

M.L.C. was responsible for project administration, conceptualisation, and investigation, writing and execution of software, formal curation, analysis, validation and visualisation of data and writing the original manuscript draft and editing.

M.H. was responsible for formal analysis and reviewing and editing draft works.

S.P.D. was responsible for supervision, resource provision, formal analysis, reviewing and editing draft versions.

All authors have read and agreed to the published version of the manuscript.

6.2 *Salmonella* Genomic Island 1 is Broadly Disseminated within Gammaproteobacteriaceae

This publication addresses Aim 3 of the thesis. Please note that figure numbers have been changed from the original publication to ensure they are consistent with the thesis. Readers will likely find viewing figures via the attached PDF copies more appropriate than the in-text figures.

6.2.1 Abstract

Salmonella genomic island 1 (SGI1) is an integrative mobilisable element that plays an important role in the capture and spread of multiple drug resistance. To date, SGI1 has been found in clinical isolates of *Salmonella enterica* serovars, *Proteus mirabilis*, *Morganella morganii*, *Acinetobacter baumannii*, *Providencia stuartii*, *Enterobacter* spp. and recently in *Escherichia coli*. SGI1 preferentially targets the 3´-end of *trmE*, a conserved gene found in the Enterobacteriaceae and among members of the Gammaproteobacteria. It is therefore hypothesised that SGI1 and SGI1-related elements (SGI1-REs) may have been acquired by diverse bacterial genera. Here, Bitsliced Genomic Signature Indexes (BIGSI) was used to screen the NCBI Sequence Read Archive (SRA) for putative SGI1-REs in Gammaproteobacteria. Novel SGI-REs were identified in diverse genera including *Cronobacter* spp., *Klebsiella* spp. and *Vibrio* spp. and in two additional isolates of *Escherichia coli*. An extensively drug resistant human clonal lineage of *Klebsiella pneumoniae* carrying an SGI1-RE in the United Kingdom and an SGI1-RE that lacks a class 1 integron were also identified. These findings provide insight into the origins of this diverse family of clinically important genomic islands and expand the knowledge of the potential host range of SGI1-REs within the Gammaproteobacteria.

6.2.2 Introduction

Salmonella genomic island 1 (SGI1) is an integrative mobilisable element that carries diverse antibiotic resistance genes, often conferring multidrug resistance (217). SGI1 was shown to be associated with globally dispersed *Salmonella enterica* serovar Typhimurium DT104 that rose to prominence in the 1980s. Since then, variants of SGI1 including some that carry genes encoding resistance to critically important antimicrobials have been identified within diverse serovars of *Salmonella enterica* and other Gammaproteobacteria including *Proteus*, *Acinetobacter*, *Morganella*, *Providencia*, *Enterobacter* and *Escherichia* (38, 39, 218-220), all of which are known human and/or animal pathogens. While experimental integration of SGI1 into genera such as *Vibrio* and *Klebsiella* has been described (41), there are no detailed reports of an SGI1-related element (SGI1-RE) in wild type strains belonging to these genera. Numerous Gram-negative genera carry the *trmE* gene, the integration site for SGI1 (40), suggesting that SGI1 may be dispersed more broadly in Gammaproteobacteria than is reported in the

literature. Here, we aimed to explore the SRA to garner a greater understanding of inter-genera distribution of this clinically important integrative genomic element.

6.2.3 Materials and Methods

Preliminary screening of SRA for SGI1-REs and subsequent genome assembly

BIGSI (216) allows for ultra-fast searching of bacterial and viral genomic data for nucleotide sequences of interest, providing access to the 457,000 whole genome sequence (WGS) datasets uploaded to SRA prior to December, 2017. These whole genome sequence datasets were screened for forty-four coding sequences sourced from SGI1 (GenBank accession No. AF261825), constituting all coding sequences from SGI1 other than those downstream of its 3` end, using the BIGSI web tool (http://www.bigsi.io/) in combination with a custom-built script available at https://github.com/maxlcummins/SGI1-REs. Retrieved accession numbers and their putative genera classifications, as designated by BIGSI via integrated Kraken (221) and Bracken (222) based analyses, were filtered to remove strains identified as *Salmonella*, *Proteus* and *Acinetobacter*; genera that have been reported previously to carry SGI1-REs. Kraken and Bracken were also used to identify genomes that were contaminated with non-host DNA. Genomic data sets were then downloaded using parallel-fastq-dump version 0.6.3 (https://github.com/rvalieris/parallel-fastq-dump) and were assembled with Shovill version 1.0.4 (https://github.com/tseeman/shovill) using default settings.

Putative identification and clustering of SGI1-RE variants

A non-redundant nucleotide database consisting of genes sourced from various SGI1-REs, available at https://github.com/maxlcummins/SGI1-REs, was generated using CD-hit version 4.8.1. Assemblies were screened for the carriage of these genes using BLASTn with a nucleotide identity threshold and coverage threshold of 90%. Strains with identical genotypic profiles, relative to these screened genes, were preliminarily considered to carry the same SGI1-RE. This analysis also informed the particular scaffolds on which genes from the backbones of SGI1-RE were localised.

Annotation and comparative analysis of SGI1-REs

For each respective strain found to carry a putative SGI1-RE, BLASTn was used to determine the location of the *trmE* and S044 genes, the former of which contains the 5' DR-L (terminal 18 bp of the *trmE* gene) and the latter of which is typically seen adjacent to the 3' DR-R (imperfect direct repeat of DR-R). This region of sequence, inclusive of the DR-R and DR-L, was then isolated and screened against publicly available SGI1 sequences deposited in the NCBI nucleotide collection using MEGABLAST. Following identification of closely related SGI1-RE sequences on NCBI, genome annotation was performed manually using Snapgene (https://www.snapgene.com) version 4.1.9.

Phylogenomic analysis and genotypic characterization of clonal subpopulations

Cases where multiple strains of the same predicted genera were found to carry identical SGI1-RE genotypes (see above) were interrogated to determine the phylogenetic and genotypic relatedness of such strains. Phylogenetic analysis was undertaken using Snippy (https://github.com/tseeman/snippy) and Gubbins (223) as previously described (224). Phylogenetic trees were generated using FastTree2 version 2.1.10 (151). Genotyping was undertaken using Abricate (https://github.com/tseeman/snippy) version 0.9.3. Trees were visualised and annotated using R version 3.6.0 with a custom script available at https://github.com/maxlcummins/SGI1-REs.

Data availability

Genomic data sets can be retrieved from National Centre for Biotechnology Information's (NCBI) Sequence Read Archive (SRA) via the following Sequence-Read Archive Accession numbers: *Cronobacter sazakii*: SRR1619558; *Escherichia coli*: ERR1617816, ERR1622475, SRR3987521, SRR4786187; *Klebsiella pneumoniae*: ERR314398, ERR314418, ERR314466, ERR314533, ERR314534, ERR314541, ERR314542, ERR314543, ERR314544, ERR314545, ERR314546, ERR486792*; Vibrio cholerae*: ERR117476, ERR180912, ERR386635, ERR386714, ERR386735, ERR386756, ERR572602, ERR572603, ERR572604, ERR572741, ERR579067, ERR579915.

Please note that we are not the submitters for these specific genomic data sets. Short read sequences utilised in this investigation are all publicly available on the Sequence

Read Archive (SRA). For specific SRA accession numbers please see Supplementary Table 1 (Additional file 6.1).

The nucleotide accession numbers pertaining to SGI1-REs identified during this investigation for KGI, PGI2-Ec-2, SGI1-Vc2CHAMA and VGI are MN708012, MN708013, MN708014 and MN708015, respectively.

6.2.4 Results and Discussion

Thirty strains belonging to the genera *Vibrio* (n=12/30), *Klebsiella* (n=12/30), *Escherichia* (n=5/30) and *Cronobacter* (n=1/30) were found to carry at least one open reading frame associated with the backbone of SGI1 (Supplementary Table 1 [Additional file 6.1]). These 30 sequences were identified after Kraken and Bracken analyses were used to exclude contaminated sequences and filtering of samples belonging to genera with multiple reports of SGI1-RE carriage (i.e. *Salmonella*, *Proteus* and *Acinetobacter*). Preliminary BLAST analysis using a non-redundant nucleotide database of SGI1-RE coding sequences indicated that 12 different SGI1-RE variants were present within the 30 isolates (data not shown). Of these, four variants could be localised to a single scaffold, while the remaining seven assembled from multiple scaffolds. Strain metadata and their associated SGI1-RE variants are listed in Supplementary Table 1 (Additional file 6.1).

Potential discovery of an ancestral variant of Salmonella genomic island 1

A schematic diagram of the SGI1-RE from Vibrio strain 2710-CN (SRA Accession No. ERR579067) that was isolated in China in 1986 is depicted in Figure 6.1A. Our analysis shows that 2710-CN carries an SGI1-RE sharing 99.97% sequence identity with the backbone of SGI1 (GenBank Accession No. AF261825). To our knowledge, this structure represents the earliest known example of an SGI1-RE. The SGI1 variant, here named VGI (Accession No. MN708015), does not harbour a class 1 integron and, given the extensive sequence identity shared with SGI1, may constitute an ancestral structure from which SGI1 and potentially other SGI1-REs may have evolved.

A recent investigation reported the identification of an SGI1-RE, SGI0 (NCBI Accession No. MG201402.1), which also lacks a class 1 integron (225). Notably, however VGI is

more closely related to both SGI1 and SGI2 than SGI0. SGI0 and VGI differ by 972 single nucleotide polymorphisms (SNPs), whereas VGI and the SGI1 backbone differ by just 6 SNPs and VGI and SGI2 backbones differ by only 136 nucleotides. It is therefore more likely that SGI1 and SGI2 are descendants of a VGI-like element than they are descendants of SGI0.

Identification of a clonal outbreak of extensively drug resistant Klebsiella pneumoniae carrying a novel SGI1-RE

Klebsiella pneumoniae strain 2485STDY5477984 (SRA Accession No. ERR314534) was found to carry a novel and complex SGI1-RE backbone (Figure 6.1B). This structure, referred to as KGI (Accession No. MN708012), can be classified into two distinct regions (R1 and R2) based on the SGI1-REs with which it shares extensive sequence identity. Spanning from DR-L to base 5984, R1 is most closely related to PGI2 (Accession No. MG201402). Secondly, R2 constitutes a region from base 6969 to 27,754, corresponding to an intragenic region of S005 and an intergenic region between *res* and S044, respectively. This portion of KGI is most closely related to SGI1-PmCA11 (Accession No. MH990673), with which it shares 99.25% nucleotide identity. It is notable that R1 and R2 are separated by a 985 bp region within S005 which shares low nucleotide identity with the corresponding region in SGI1 and PGI2. Analysis using BLASTn indicates that this region matches part of the *traN* gene of PGI1 (Accession No. KJ411925) with 87% nucleotide identity, despite an alignment of the full $traN_{PGI1}$ and $traN_{KGI}$ coding sequences revealing a shared nucleotide identity of just 77% and a coverage of just 81%. KGI also features an S044 gene with relatively low nucleotide identity to SGI1 and PGI2 (82% and 83%, respectively), which instead best matches $S044_{PGI1}$ – a gene with which it shares 99% sequence identity. This data indicates that KGI likely has a complex evolutionary history and has a mosaic structure resulting from homologous recombination events involving different SGI1-REs.

It is notable that a further 10 strains from the same Bioproject (PRJEB1271), several of which were listed as being sourced from humans in the United Kingdom (226), also carried an element identical to KGI. Subsequent analysis determined that these strains are also genotypically extensively drug resistant (XDR), exhibited nearly identical genotypic resistance profiles (two samples did not carry *aphA1*) and differed from one

another by between zero and five SNPs (Figure 6.2). These strains, therefore, appear to constitute a clonal outbreak of an SGI1-RE carrying XDR species that are a frequent cause of nosocomial infections.

Novel variants and hosts of SGI1-related elements

A second SGI1-RE, here named Vc2CHAMA, was identified in *Vibrio cholerae* strain 59-sc-2011-11-25T13:51:24Z-1311470 (SRA Accession No. ERR117476). Vc2CHAMA shares 99.99% sequence identity with SGI1 Pm2CHAMA (Accession No. MF372716.1) and carries a single In*4*-type integron (Figure 6.3A). Both structures carry a *dfrA15* gene cassette and include a unique 34 bp deletion at the start of S044 caused by IS*6100* – a deletion not otherwise detected by MegaBLAST in the NCBI nucleotide database and one that indicates that SGI1 Pm2CHAMA and Vc2CHAMA share a recent common ancestor. SGI1 Pm2CHAMA and SGI1 Vc2CHAMA differ however in that the former carries a complex resistance region containing two integrons, multiple copies of IS elements, including IS*26,* and an additional five antimicrobial resistance genes (*aadA1, aph1a, strA, strB and bla*$_{\text{CARB-4}}$). It is possible that SGI1 Vc2CHAMA may have arisen from a homologous recombination event between the two integron integrase genes in Pm2CHAMA. Alternatively, Pm2CHAMA arose from a Vc2CHAMA-like structure via homologous recombination event or via an IS*26*-mediated event introducing the upstream integron and other associated elements of the complex resistance locus.

Additionally, a PGI2 variant, here named PGI2-Ec-1 was identified in *Escherichia coli* strain SCP17-71-2 (SRA accession No. ERR1617816). This strain was identified in a study investigating the phylogenetic relatedness of 1,798 ESBL-producing Enterobacteriaeceae from a single hospital in the Netherlands (227). PGI2-Ec-1 exhibits 99.28% sequence identity (99% sequence coverage) following BLASTn alignment with PGI2 (Accession No. MG201402) (Figure 6.3B). PGI2-Ec-2 carries a single In*4* type integron containing *dfrA5*. In contrast, PGI2 carries a complex and mosaic resistance locus featuring two integrons, numerous copies of IS*26* and other insertion sequences. Similarly, with SGI1-Vc2CHAMA, PGI2 and PGI2-Ec-1 could have evolved in either direction via integron integrase associated homologous recombination events or by IS*26*-mediated insertion. While a recent study of ours reported the first identification of an SGI1 variant within *E. coli,* PGI2-Ec-2 constitutes, to our knowledge, the first report

of a PGI2 variant within the *Escherichia* genus. This sequence type is known to cause extraintestinal infections in both humans and animals and coincidently, is also the same sequence type of the strain harbouring SGI1-B-Ec1 that we recently reported (220).

Additionally, a portion of the backbone of an SGI1-RE within a strain of *Cronobacter* called CFSAN019572 was identified by our analyses. However, a scaffold break thought to be caused by the insertion of an IS*21* family element into *rep* prevented assembly of the complete sequence. The specific aforementioned IS element could not be determined, however, as it was not assembled to a single scaffold. Notably, the integron associated inverted repeats IRi and IRt are not flanked by 5 bp direct repeats, indicating that this locus is unlikely to have resulted from an integration event (228, 229) and may instead be a remnant of an earlier homologous recombination event. Preliminary analysis revealed that the structure is flanked by a DR-L and DR-R and appears to have integrated into the terminal 18 bp of the *trmE* gene. It was also noted that the *intI1* gene in this region differed from that of SGI1/PGI2 by 18 SNPs. Only 1 hit within the NCBI nucleotide database (Accession No. LT629801) contained an *intI1* sequence with 100% sequence homology – a strain of *Pseudomonas rhodesiae* which also carried the same chlorine dismutase, *tniC* and *tniQ* genes as the *Cronobacter* integron under analysis, indicating that they share a common origin. An additional six SGI1-RE variants were also identified (Supplementary Table 1 [Additional file 6.1])), however they will not be described in detail due to an inability to determine their complete sequences. Further investigation into these strains, perhaps using long-read sequencing or PCR linkage of SGI1-RE associated scaffolds, would likely reveal novel SGI1-RE variants.

In conclusion, these findings provide insight into the potential evolutionary origins of SGI1 and provide evidence for a greater SGI1 host range than previously recognised. They also highlight that SGI1 and SGI1-REs are part of a large and diverse family of integrative genomic elements whose structures are shaped by a wide range of events including insertion, deletion and homologous recombination. Once these islands are captured by members of the Gammaproteobacteriaceae, they may come under selection in environments affected by anthropogenic pollutants including antimicrobials, heavy metals and pharmaceutical residues. As such it is unsurprising that SGI1-RE can readily acquire multiple antibiotic resistance genes, including genes

encoding resistance to clinically important extended-spectrum beta-lactams (218). These SGI1-REs carrying AMR gene cargo can then disseminate more broadly as they provide their host a fitness advantage in the presence of antimicrobials. Further investigations into the host range of SGI1-REs and more directed searches of these elements within clinically relevant Gram-negative species are needed. More research is also needed to elucidate the phylogenetic relatedness of SGI1-REs. Comparative genomic analysis between SGI1-REs lacking resistance regions, such as those presented in this research and those identified within *Shewanella*, would likely prove informative in this regard (230).

To date, the focus has largely been on the characterisation of SGI1 and SGI1-REs in pathogens. Our studies suggest that further efforts are needed to determine the presence of SGI1-REs in commensal flora circulating in humans, food animals and the environment more broadly. Additionally, screening for SGI1-RE backbones, rather than phenotypes associated with SGI1-REs, can provide valuable insight into the spread of these elements within Gammaproteobacteriaceae. While the subset of isolates analysed comprises 456,000 genomic datasets uploaded prior to 2017, there are more than 900,000 genomic datasets that are currently available in the SRA. This, however, would require a reindexing of SRA by the tool's developers. Nonetheless, screening of these additional datasets will likely identify new SGI-REs, enhancing our understanding of the spread of these diverse and clinically significant genomic islands throughout parts of the bacterial Kingdom.

6.2.5 Supplementary Materials

The following is available online at https://msphere.asm.org/content/msph/4/3/e00169-19/DC1/embed/inline-supplementary-material-1.docx?download=true

6.2.6 Author Contributions

M.L.C. project administration, conceptualisation, and investigation, writing and execution of software, formal curation, analysis, validation and visualisation of data and writing the original manuscript draft and editing. M.H. formal analysis and reviewing and editing draft works. S.P.D. supervision, resource provision, formal analysis, reviewing

and editing draft versions. All authors have read and agreed to the published version of the manuscript.

6.2.7 Funding

This project was partly funded by the Australian Centre for Genomic Epidemiological Microbiology (AusGEM), a collaborative partnership between the NSW Department of Primary Industries and the University of Technology Sydney. M.C. is a recipient of an Australian Government Research Training Program Scholarship. M.H. is supported by the University of Technology Sydney Chancellor's Postdoctoral Research Fellowship (CPDRF PRO17-4005). These results are helping to inform the OUTBREAK decision support system funded by the Medical Research Future Fund Frontier Health and Medical Research Program.

6.2.8 Acknowledgments

We would like to extend our gratitude to all nucleotide sequence submitters for their public sharing of genomic data sets, through which this manuscript was able to be written.

6.2.9 Conflicts of Interest

The authors declare no conflict of interest.

6.2.10 Figures

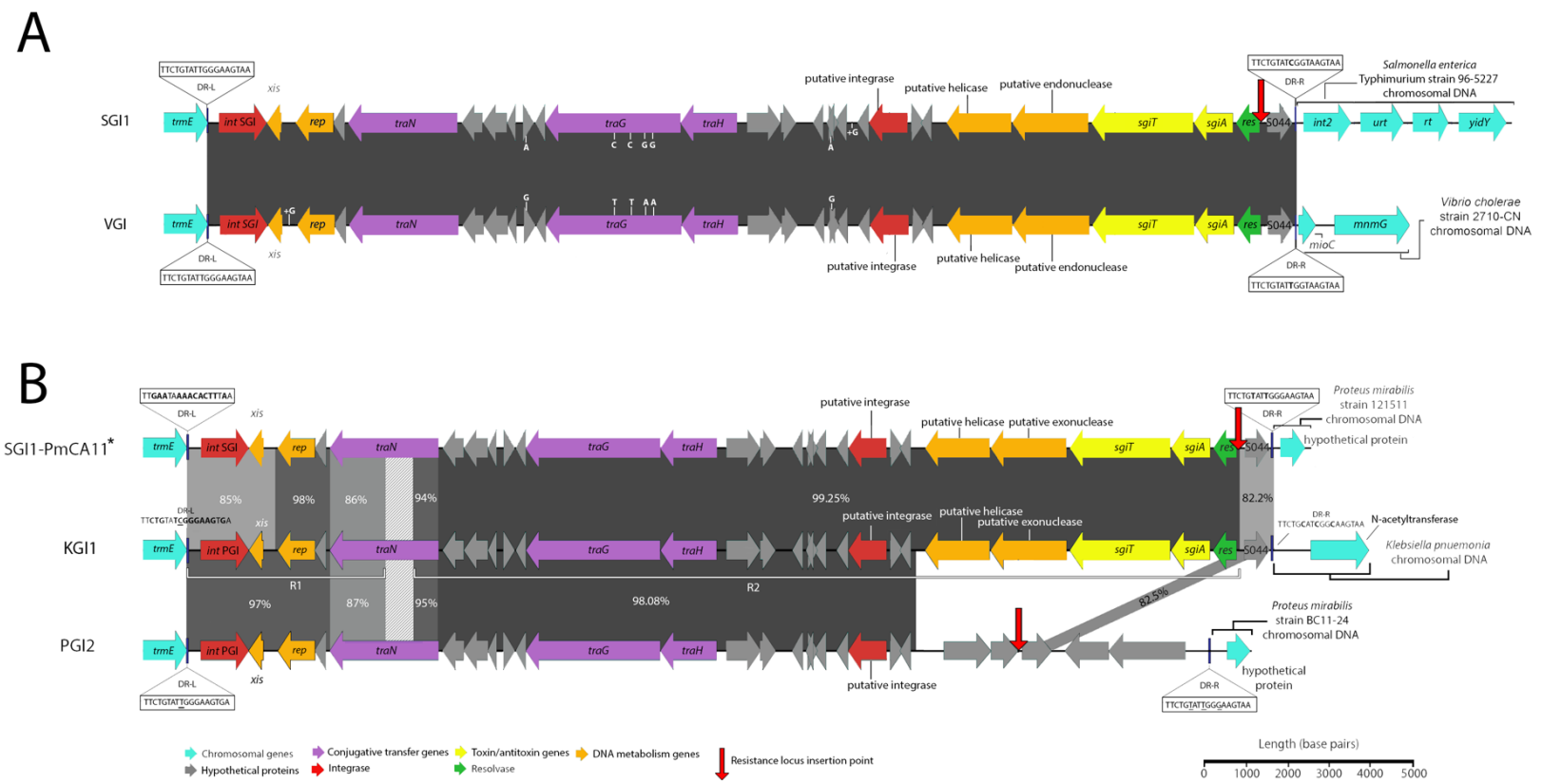

Figure 6.1 - Schematic representation of SGI1-REs identified in this investigation that are void of integrons. Scale is approximate. Closely related SGI1-REs from public databases are shown atop and below to allow visualization of the nucleotide identity across the novel elements and the references. Nucleotide identities of *trmE* genes (not shown) did not exceed greater than 75% for VGI or KGI in comparison with the *trmE* genes of other sequences represented in this schematic. (A) – A comparison of SGI1 (Accession No: AF261825) shown atop VGI (Accession No. MN708015). Identities and locations of single nucleotide polymorphisms are indicated atop vertical lines at the top and bottom of either element, while nucleotide insertions are indicated in the same fashion but denoted with a plus symbol (+). Genetic context inferred from *S.* Typhimurium DT104 (GenBank accession number HF937208). (B) A comparison of KGI (Accession No. MN708012), PGI2 (Accession No. MG201402) and SGI1-PmCA11 (Accession No. MH990673). Due to the lower nucleotide identity matching between these structures the locations and identities of mismatches and indels are not indicated for ease of interpretation. SGI1-PmCA11*: genetic context inferred from Xiao et al (2019).

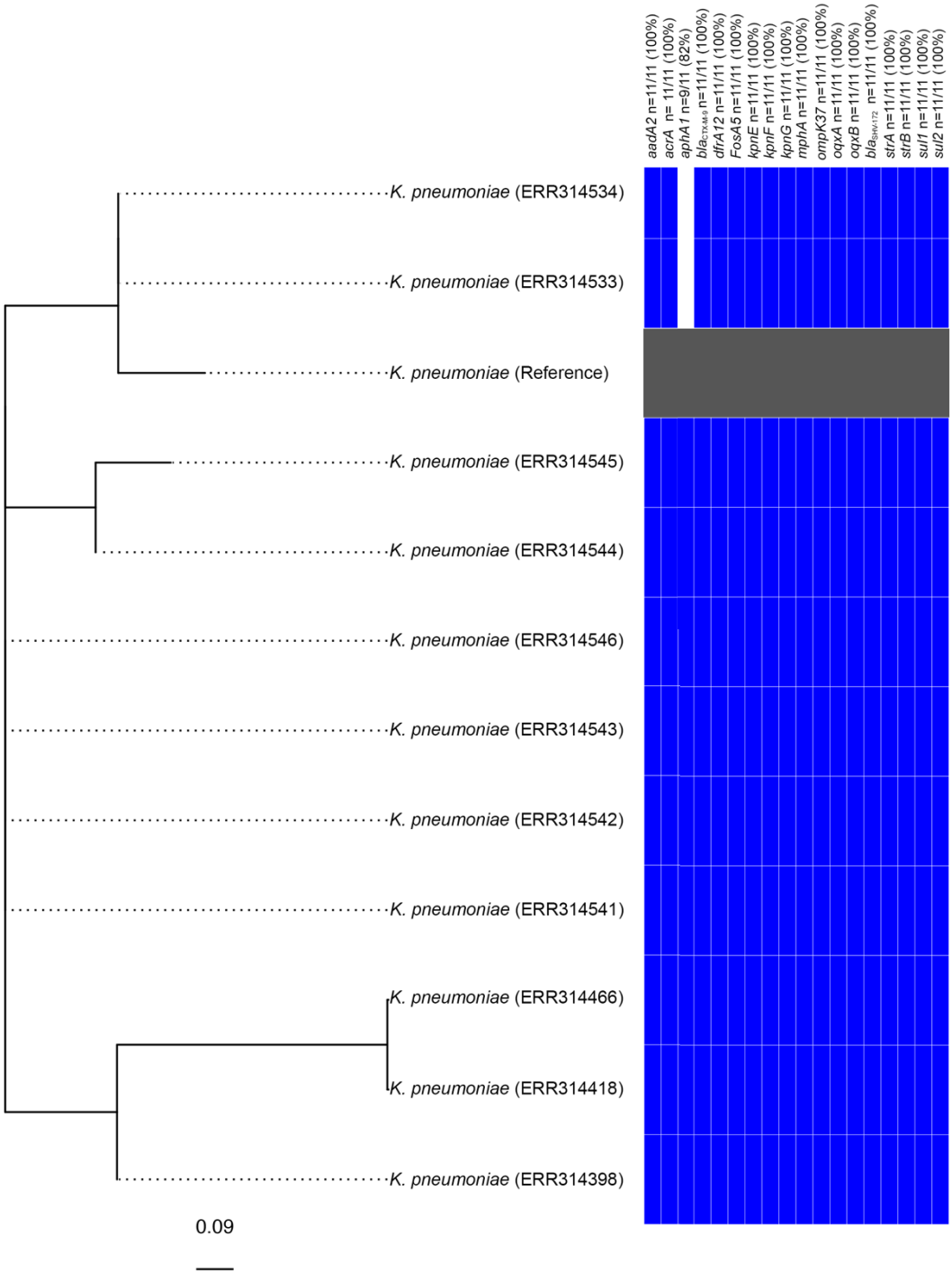

Figure 6.2 - Phylogenetic analysis of *Klebsiella* strains found to carry KGI. The tree is unrooted and the reference strain utilised corresponds to an assembly of *Klebsiella pneumoniae* strain 2485STDY5477984 (SRA Accession No. ERR314534)

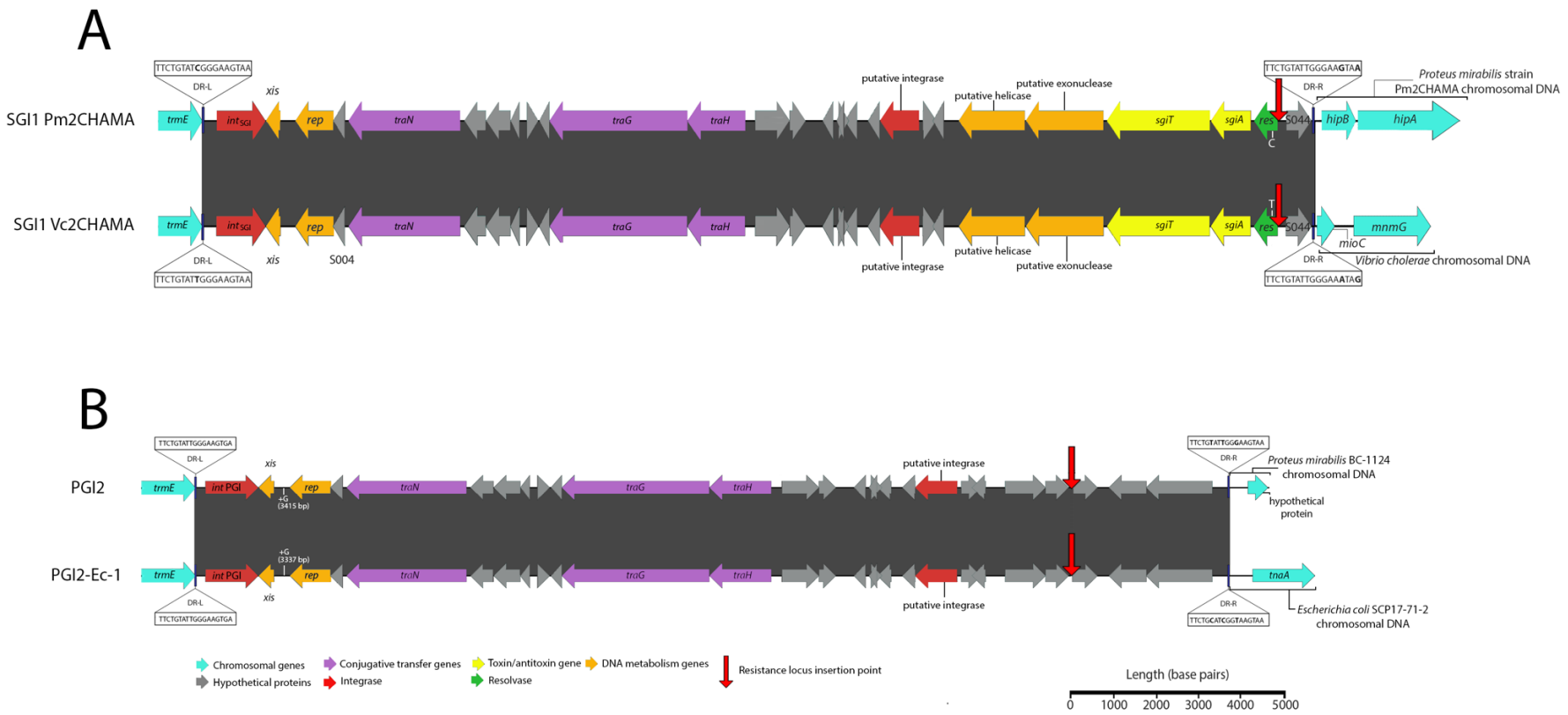

Figure 6.3 - Schematic depicting SGI1-REs identified in this investigation. Scale is approximate. Closely related SGI1-REs from public databases are shown to allow visualization of the nucleotide identity across the novel elements and the associated references. Identities and locations of single nucleotide polymorphisms are indicated atop vertical lines at the top and bottom of either element, while nucleotide insertions are indicated in the same fashion but denoted with a plus symbol (+). Nucleotide identities of *trmE* genes (not shown) did not exceed greater than 75% for SGI1 Pm2CHAMA or PGI2-Ec-1 in comparison with the *trmE* genes of other sequences represented in this schematic. (A) – A comparison of SGI1 Pm2CHAMA (Accession No: MF372716) shown atop SGI1 Vc2CHAMA (Accession No. MN708014). (B) A comparison of PGI2 (Accession No. MG201402) and PGI2-Ec-1 (Accession No. MN708013)

Chapter 7: General Discussion and Future Directions

Please note: Given that each individual chapter of this thesis discusses the implications of their respective results, this section will provide a brief discussion of the findings presented throughout this thesis, their limitations and the future directions of this research. It will then close with some final thoughts on future One Health approaches to genomic investigations.

7.1.1 APEC as a potential source of AMR, zoonoses and virulence plasmids

Data presented in Chapter 4 indicates that Australian poultry are not commonly colonised by CIA-resistant microorganisms, and this trend is also seen in Australian piggeries (28, 215). In many countries the same cannot be said (176-178), and the success of Australia in this regard is likely attributable to our geographical isolation, strict biosecurity programs and prudent agricultural usage of CIAs. However, should CIA-resistant microorganisms become established and spread within Australian food-animal production systems the results could be devastating, highlighting the need for widespread and ongoing AMR surveillance programs.

For example, the increased prevalence in colistin resistance documented in clinical settings in China, Italy and Spain (6, 231, 232) is likely attributable to historic usage of colistin in food-animal systems in these countries. China's 2016 colistin consumption as agricultural growth promoters was 8000 tonnes, compared to less than 500 tonnes used across the entirety of the European Union (EU) in 2013 (233). Notably, Italy and Spain reported higher use of colistin per food-producing animal than many other EU countries combined (234), and there are several reports of high carriage rates of *mcr-1* in food-animal operations in these countries. Overseas, poultry operations have also been shown to house colistin resistant *E. coli* (176-178).

Aside from serving as a source of AMR, the risk posed by APEC, and poultry-associated *E. coli* more broadly, is yet to be fully characterised. Given that there is evidence APEC may threaten human health their populations warrant further scrutiny and monitoring. Phylogenetic overlap has been observed between APEC and pandemic lineages of ExPEC, in particular ST95, one of the most frequent STs of *E. coli* recovered from cases of blood sepsis globally (169). Recent studies have highlighted the similarities between

infectious ST95 strains from both poultry and humans (235, 236), however the between-source phylogenomic and genomic similarities of members of this lineage requires further investigation. Such investigations should utilise high resolution phylogenetic approaches involving combinations of pangenomic, SNP-based and core-genome MLST methodologies. Detection of clonal overlap between strains of poultry-associated *E. coli* and human ExPEC may indicate transfer events between these two reservoirs and would support the need for an expanded screening regime of avian faecal *E. coli*, APEC, retail poultry meat-associated *E. coli,* human commensal *E. coli* and human ExPEC to further elucidate the zoonotic potential of poultry-associated *E. coli*, especially APEC. The impact of the use of poultry manure on the contamination of fresh produce has not been investigated using WGS, however studies in Germany (51) suggest that animal waste is an underestimated and important source of MDR Enterobacteriaceae in fresh produce.

Even if links to human health were to be discounted, APEC will continue to be a major burden on poultry operations at home and abroad and similarly require genomic surveillance to ensure their sustainability. Evidence indicates that particular lineages of *E. coli* are disproportionately associated with APEC infections, such as ST117. Research is required to elucidate the genetic underpinnings of their success within this niche to try to reduce their disease burden on poultry operations. It is probable that this sequence type has characteristics that enhance its ability to colonise the poultry gut, leading to its high prevalence in these environments. It is also known that some of the fitness factors involved in gastrointestinal colonisation are linked to extraintestinal infections in both poultry and in humans (77, 138). While there is some evidence of some protective effect of one commercially available APEC vaccine, other efforts for APEC and human ExPEC vaccines have been largely unsuccessful(80, 194, 197, 237). Genome wide association studies may prove useful in identifying previously unknown VAGs common that underpin APEC and ExPEC pathogenesis that may hold value as potential vaccine targets for human and/or animal populations.

Analysis of Australian MDR APEC identified a variety of *repA* genes, providing evidence that these populations carry a diverse array of plasmids. Most prevalent among this APEC population, however, were plasmid replicons of incompatibility type F that are often associated with ColV and ColV-like virulence plasmids. Given the high prevalence

of these plasmids within APEC populations and the known associations of their genetic cargo with virulence in both avian and human hosts, their abundance within poultry reservoirs is problematic. Screening of the poultry gut microbiome will be important in this regard as it is thought to be the primary reservoir of APEC and therefore also a reservoir of ColV-like plasmids (93). Addressing this issue may reap significant benefits in animal health and economic productivity of food-animal production through reducing the disease burden on poultry operations. This could be achieved through the usage of a probiotic and/or prebiotic formula to attempt to shift the poultry microbiome to one with lower VAG carriage, though this may be made difficult by the inherent commensal fitness attributes also thought to be associated with these genes.

The impact of *E. coli* that carry ColV plasmids on human health also warrants more attention, as no large scale studies have investigated the carriage of ColV-like virulence plasmids among human ExPEC strains, nor on *E. coli* from other sources. Small scale analysis of *E. coli* causing UTIs and BSI from Enterobase suggest that between 6-12% of *E. coli* carry ColV and closely related plasmids (data unpublished), according to the Liu criteria of carriage of ColV-like plasmids (238). An expansion of such analysis is critical for complete assessment of the potential threat of ColV-like virulence plasmids. This is particularly important given there are increasing reports of hybrid virulence-resistance ColV and ColV-like plasmids being reported in the literature (101, 239), the transfer of which can potentially confer both extraintestinal virulence and AMR to an otherwise avirulent and antimicrobial susceptible strain of *E. coli*.

7.1.2 SGI1-REs and their potential spread within Australian poultry

As reported in Chapter 5 we identified a strain of *E. coli* carrying a variant of SGI1. This novel finding demonstrates the diversity of MGEs present within poultry environments and highlights the need for more extensive analysis of poultry-associated *E. coli* and the poultry microbiome more broadly. Future investigations should involve widespread screening of poultry-associated flora for SGI1-REs, particularly those that might house genes conferring resistance to CIAs.

The strain housing SGI1-B-Ec1 was ST117 – a globally disseminated APEC lineage, and emerging human ExPEC, which seems exceptionally fit within the poultry gut, being routinely associated with poultry meats, poultry faeces and poultry pathogens (73, 74,

99, 161). Should a strain of *E. coli* carrying a CIA resistance associated SGI1-RE also be a member of a 'poultry adept' ST, particularly if it entered into higher tiers of the poultry breeding pyramid, it could rapidly disseminate throughout poultry operations around Australia and make its way into human populations and other environments via a range of mechanisms. In such a scenario, the most important pathways are likely to include consumption of poultry products and widespread application of poultry manure.

7.2 Limitations

7.2.1 Aim 1

Determine the phylogenetic characteristics, virulence and AMR associated gene content of a geographically-diverse collection of Australian APEC

Aim 1 of the thesis was addressed in Chapter 4, which investigated a collection of 95 Australian APEC which were sampled based on the carriage of *intI1,* a proxy for MDR.

Some of the primary findings of this chapter were that Australian MDR APEC are:

- Not a major reservoir of genes conferring resistance to CIAs
- Rich in ExPEC/APEC virulence gene content
- Phylogenetically diverse but most frequently of ST117 and CC350

As Chapter 4 has already discussed the implications of these findings, this section will instead address the limitations of the study.

Firstly, the collection of APEC was not produced using a stringent sampling protocol. Hosts from which the strains were isolated consist primarily of diseased chickens, but some strains were isolated from turkeys and ducks (Additional file 4.1). While not yet conclusively demonstrated, evidence suggests that some lineages of *E. coli* have a degree of host-specificity, and therefore the phylogenetic characteristics of the APEC collection are likely to partially reflection the hosts from which they originate (240, 241). For example, one of the primary STs identified within the collection was identified exclusively within turkeys, ST429. It is possible that ST429 may have an enhanced capacity to cause APEC infections in turkeys and may be a lineage of concern for those in this part of the poultry industry. As reported in Chapter 4 and elsewhere, ST429 is

known to be associated with turkeys and carries a high number of VAGs (215, 242). Splitting the collection into subpopulations based on host species may have been useful, however this was not undertaken due to the low representation of particular host-types within the collection. Ideally, such an investigation would span multiple poultry operations from diverse geographic locations and perform WGS with many hundreds or thousands of isolates.

Metadata pertaining to the historic usage of antimicrobials at each poultry operation, perhaps through a standardised questionnaire, would have been useful for the purposes of this investigation. Differences in usage of antimicrobials at particular farms may have influenced the resistance phenotypes observed within the collection. For example certain antimicrobials used in broiler chickens such as cannot be used in layer chickens as these antimicrobials contaminate egg products (243), however the number of APEC sourced layer chickens within the collection prevented statistically useful comparisons of APEC from different types of farms.

One advantage of the collection was that it covered a broad geographical area, consisting of strains sourced from four different states around Australia. However, these states were not equally represented within the collection, with Victoria and New South Wales (NSW) contributing the majority of strains. This is in large part attributable to the veterinary practice responsible for composing the collection being located in Victoria and predominantly servicing clients in Victoria and NSW.

Regardless of the shortcomings of the composition of the APEC collection under investigation in Chapter 4, prior to the publishing of the data presented in this thesis there was very little known about Australian APEC populations at a genomic level. The genomic datasets generated during this investigation are therefore of value in One Health investigations comparing APEC and other populations of *E. coli*, such as human ExPEC, and to provide insight into the virulence and AMR characteristics of Australian MDR APEC.

7.2.2 Aim 2

Characterise mobile genetic elements associated with Australian MDR APEC

Knowledge on the dominant and potentially emerging MGEs present within a given environment is necessary to assess the threat a given environment may pose in relation to AMR and infectious diseases. This former point was primarily addressed in Chapter 4, while the latter point was addressed in Chapter 5.

Within Chapter 4 we provided evidence that Australian MDR APEC are rich in VAGs associated with ColV-like plasmids and frequently carried *repA* genes that also indicate the presence of such plasmids. This was bolstered by the read mapping approach demonstrated in Figure 4.4, which bolsters the evidence for the presence of such plasmids within the majority of APEC strains under investigation. It should be noted, however, that there are limitations to this approach.

Assembly of short-read sequence data inherently produces fragmented genomes, particularly in MDR bacteria as AMR genes tend to be associated with carriage of IS elements. This is due to multiple copy number and their length being longer than that of Illumina short-reads, all of which serve to decrease assembly contiguity and therefore reduce the linkage of genomic loci (112). While some APEC/ExPEC VAGs are commonly seen on ColV-like virulence plasmids, some of these same VAGs are also known to exist in chromosomal contexts (86). In short, we are unable to conclusively demonstrate the linkage of the ColV-associated VAGs to a ColV-like plasmid with short-read sequence data alone. However, based on previous genomic characterisation of ColV and ColV-like plasmids and the known association of APEC with such plasmid types, it is highly likely that these VAGs do indeed reside in such a genomic context.

In contrast, identification of an SGI1-RE within AVC96 during the analysis of the collection reported in Chapter 5 provided an opportunity to undertake long-read sequence analysis on one of these Australian MDR APEC strains. While short-read sequencing alone was unable to accurately determine the genomic context of SGI1-B-Ec1, hybrid assembly with long-reads revealed its complete nucleotide sequence as well as its chromosomal context.

Given that Chapter 5 is observational in nature there are fewer limitations of the study that can be raised than for other chapters. However, we reported that its potential spread within Australian food-production systems is a point of concern, particularly if such SGI1-REs confer MDR and, worse still, carry genes conferring resistance to CIAs. Future studies should undertake widespread screening of food-animal associated flora to determine the potential presence of SGI1-REs, particularly within major human pathogenic genera such as *Salmonella*, *Escherichia* and *Proteus.* It would also be insightful to determine the *in-trans* conjugative capacity of SGI1-B-Ec1 and the biological consequences of the interruption of S023 by the integration of ISEc43.

7.2.3 Aim 3

Assess the spread of Salmonella Genomic Island 1 and SGI1-related elements throughout publicly available genomic datasets

In Chapter 5 we reported on the first identification of an SGI1-RE within an *E. coli* strain, AVC96. Subsequently, as reported in Chapter 5, we were interested in further elucidating the spread of SGI1-REs within Gammaproteobacteriaceae.

The primary findings of Chapter 6 were as follows:

- *Cronobacter*, *Klebsiella* and *Vibrio* species carry SGI1-REs
- A clonal outbreak of XDR Klebsiella *pneumoniae* carrying a complex and likely hybrid SGI1-RE may have taken place in the United Kingdom between around 2011
- A strain of *Vibrio cholerae* from the 1980s houses an SGI1-RE that resembles an ancestral variant of SGI1

While this study provided further insight into the spread and potential evolutionary history of SGI1 and SGI1-REs, it has several limitations. Firstly, online sequence databases tend to be heavily biased towards clinically important microorganisms (244), such as those that exhibit AMR and pathogenicity. For example, ST131 constitutes a global pandemic lineage of ExPEC that frequently carries genes conferring resistance to ESBLs and SNPs conferring fluoroquinolone resistance (71). Enterobase is a large online genome sequence database and analysis pipeline that can be leveraged for the use in genomic epidemiological investigations (75). At time of writing, it contains

approximately 130,000 strains of *E. coli,* and 8,446 of constitute ST131 strains. ST11 (typically reported to be serotype O157:H7/H-), which is commonly associated with shiga-toxin production and is regularly responsible for outbreaks of diarrhoeal disease, is also likely over-represented, with 12,757 strains of this sequence type present in enterobase. Together, these two STs therefore constitute more than 15% of the Enterobase collection. This trend is likely a feature of other sequence databases as well, as researchers chasing impactful scientific narratives have a greater incentive to study microorganisms of clinical significance rather than those associated with commensal and environmental niches. In short, a greater diversity of SGI1-REs are likely present within a wider number of genera than can otherwise be gleaned from public databases due to their inherent biases.

Another limitation is present in BIGSI, the tool used to screen the Sequence Read Archive for SGI1-REs, as it is only indexed to December 2016 (216). Costs of WGS have dropped by many orders of magnitude in recent decades and there is an inverse relationship between the cost of sequencing and the number of genomes submitted to public databases. For example, within the NCBI nucleotide database the number of whole genome sequence datasets has tripled since December 2016. It is therefore likely that many additional SGI1-RE carrying genera will be identified within this newer volume of sequence data. The authors of BIGSI are reportedly reindexing the SRA database such that all sequence data generated from December 2016 onwards will be accessible via the BIGSI interface (Personal correspondence). It is highly likely that undertaking the same analytical approach upon the completion of these changes identifies many more novel SGI1-REs in a broad array of host genera that have yet to be reported to carry such structures.

7.4 Future directions

One of the most important approaches to minimising the impacts of infectious diseases and AMR will involve the undertaking of massive One Health genomic epidemiological investigations. Such investigations should involve sampling of diverse bacterial populations from a variety of sources across time and space.

Efforts should be made to minimise sampling biases incurred by screening for particular resistance phenotypes and genotypes. For example, all but three SGI1-REs reported in the literature carry AMR gene cargo (225). There is little evidence for ancestral variants of SGI1-REs from which more recent AMR-associated elements have evolved, limiting our understanding of their evolution and also their spread throughout Gammaproteobacteriaceae. A better understanding of the distribution and evolutionary history of mobile genetic elements may provide valuable insights that shape our response to the global AMR crisis.

Similarly, there is an overrepresentation in the literature of strains resistant to CIAs, such as is the case for *E. coli* ST131 in studies on human ExPEC. These biases obscure the ability of genomic epidemiologists to detect emerging pathogens and misappropriates disease burden to particular lineages (89). While it is true that strains with particular resistance genotypes and/or phenotypes may be of greater clinical significance and prevalence than others, comparisons of collections that were composed using different AMR selection criteria greatly reduces the conclusions that can be drawn from these studies. Collecting metadata on sampling protocols associated with genomic datasets may be useful in this regard.

There is also an overrepresentation of pathogenic strains in public databases. Ideally, future genomic epidemiological investigations will involve analysis not only of *E. coli* but also other pathogens of clinical significance (such as ESKAPE pathogens) and also other bacteria able to share genetic material with these pathogens. Whole genome sequencing studies of commensal *E. coli* populations is also severely lacking; antimicrobial resistance is not restricted to pathogens, and determination of carriage rates of genes encoding CIA-resistance in commensal *E. coli* from healthy individuals is urgently needed. Indeed, studies indicate travellers returning from travel to many Asian and South American countries become colonised – sometimes for periods up to 1 year - with *E. coli* carrying genes encoding resistance to CIAs (245-247).

In absence of a capacity to definitively classify *E. coli* pathotypes genotypically, isolation of individual strains will remain a critical to research identifying potential reservoirs of zoonotic bacteria. Despite the cost of DNA sequencing falling rapidly, the scale of sequencing required to interrogate multiple genera across numerous environments

throughout time and space may be cost-prohibitive. Therefore, given the widespread presence of *E. coli* in human, food-animal and natural environments, its tendency to house broad-host AMR conferring plasmids, and its prominence as a human and animal pathogen, *E. coli* makes for a suitable candidate for such a One Health investigation.

However, metagenomic sequencing and assembly serve as an increasingly promising technological platform for One Health investigations. It is important to note, though, that as it involves extensive shearing of DNA, and the inherent complexity of metagenomic samples, its ability to associate specific mobile genetic elements with particular host species is limited. This is of consequence as plasmids and their associated gene cargo are a primary means of dissemination of AMR and also, in the case of ExPEC, a primary means of dissemination of VAGs. In this way, the future of infectious diseases and AMR will be increasingly shaped by the role of plasmids. Other technologies such as Hi-C may prove fruitful in this regard. Hi-C is a method that, cross-links proximal genomic structures prior to their sequencing, allowing plasmids and other MGEs to be assigned to hosts in complex bacterial communities. An alternative strategy may involve focusing on so called “metaplasmidomic” analyses whereby researchers may attempt to isolate, purify and sequence the plasmids from a mixed microbial community. In terms of tracking and monitoring of elements important in AMR evolution and dissemination, metagenomic and metaplasmidomic approaches cast a much wider net than those involving the WGS of single strains. Metagenomic sequencing and assembly platforms are improving rapidly and will likely play increasingly important roles in genomic epidemiology in years to come.

Ultimately, transfer of microorganisms between agricultural, clinical and natural environments is a major threat to both human and animal health, as a subset of these microbial consortia: i) have a capacity to cause anthropozoonotic (animal to human), zooanthroponotic (human to animal) or sapronotic (environment to human) infections, and ii) house virulence or AMR associated genetic cargo which can be shared with other microbes to enhance their ability to cause disease and/or resist antimicrobial therapeutics. Rapid and global expansion of large scale One Health genomic epidemiological investigations, along with collaborative data sharing and analysis, will

likely prove critical in our efforts to predict and respond to the emergence and spread of AMR and infectious diseases.

Appendix

Supplementary File 5.1 - Materials and Methods

Sample collection

AVC96 was collected in 2009 from a 26-week old diseased commercial chicken that suffered from a presumptive avian pathogenic *E. coli* infection. The sample was cultured on sheep blood and MacConkey agar as well as subjected to a routine PCR (78) to enable its identification as putative APEC. It was subsequently store at -80°C in 20% glycerol or on Protect (Thermo-Fisher) beads.

DNA isolation and Illumina short-read sequencing

After overnight culturing in 5 mL of Lysogeny broth (LB), genomic DNA was extracted using the ISOLATE II Genomic DNA Kit (Bioline) following the manufacturer's instructions and stored at -20°C. Nextera® DNA Library Preparation kits were used to generate 150-bp paired end reads from 0.5ng of template DNA before whole genome sequencing using an Illumina HiSeq® 2500. FastQC version 0.11.5 was used to quality assess read quality.

High molecular weight DNA isolation for long read sequencing

Genomic DNA for use long read sequencing was isolated using a custom protocol combining the protocols of Tillett and Neilan (113) and the QIAGEN© Blood and cell culture DNA Midi kit.

Firstly, LB was inoculated and cultured overnight. A 6 mL aliquot was then pelleted at 8000rpm for 6 mins before cells were resuspended in 50 uL TER (TE buffer with 200ug/mL RNAse A). 1 mL of xanthogenate-SDS buffer (113) was added to the resuspended cells which were incubated at 50°C for 2 hours or until complete lysis was observed.

Tubes were vortexed for 10 seconds before being put on ice for 30 minutes and spun at 4°C at 14000rpm for 10 minutes before the supernatant was transferred to a new tube. The DNA was then precipitated with a 1:1 ratio of room temperature isopropanol with gentle agitation. Tubes were then spun at 4°C at 14000rpm for 10 minutes before pouring off of isopropanol, pellets were washed in 1 mL of 70% ethanol and the centrifugation was repeated. Ethanol was then poured off and the tubes were allowed to air dry in a laminar flow hood. The pellet was then resuspended in 500uL buffer B1 from the QIAGEN© kit and left at 50°C overnight.

The next morning tube contents (including any remaining pellet) were transferred to a 15 mL falcon tube and an additional 3 mL of buffer B1 was added. The protocol was subsequently carried out from step five in the Qiagen Genomic Manual, however note that here QIAGEN© Proteinase K was added but RNAse A, typically required by the protocol, was not included as they were added during the lysis step.

Single Molecule Real-Time (SMRT) cell long-read sequencing

Long read sequencing was undertaken at the Ramaciotti Centre for Genomics using a Pacific Biosciences RSII sequencer with P6-C4 chemistry on a SMRT Cell.

Long-read sequence pre-processing and assembly

Long reads were filtered using filtlong with the following command:

filtlong -1 AVC96_R1.fastq.gz -2 AVC96_R2.fastq.gz --min_length 1000 --keep_percent 90 --target_bases 500000000 A96_NovemberPacBio.fastq.gz | gzip > AVC96_filtered.fastq.gz

Subsequently, filtered short reads and long reads were combined in a hybrid assembly using Unicycler as follows:

unicycler -1 AVC96_R1.fastq.gz -2 AVC96_R2.fastq.gz -l AVC96_filtered.fastq.gz -o AVC96_hybrid

EasyFig

The coordinates of the SGI1-B-Ec1 structure from the hybrid assembly of AVC96 sequence reads were identified using ISfinder (https://isfinder.biotoul.fr) and Megablast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with SGI1B (Accession No.:KU987430) as a reference sequence. The 35,728 kb sequence of SGI1-B-EC1 (bases 4,299,832-4,335,559) was extracted from a 4,886,273 bp long scaffold from the assembly saved as a separate fasta file using Snapgene (www.snapgene.com).

This resulting fasta file was fed into Easyfig to produce a bitmap image of the homology between SGI1-B-Ec1 and SGI1B:

Easyfig.py -o figure.bmp SGI1-B-EC1.fasta SGI1B.fasta

Read-mapping

Short reads were mapped to the SGI1-B-EC1 sequence produced above to check for read depth across the element with Burrows-Wheeler Aligner using the following commands:

bwa mem -t16 -MY SGI1-B-EC1.fasta AVC96_R1.fastq.gz AVC96_R2.fastq.gz | samtools view -ubS -F 0x904 - | samtools sort -@8 -T SGI1-B-EC1.fasta - -o AVC96_SGI1-B-EC1.bam

samtools depth AVC96_SGI1-B-EC1.bam > AVC96_SGI1-B-EC1_short-read_depth.txt

To facilitate visualization of the read alignment, the following command was used before loading the bam and associated reference sequence in Tablet to confirm the bridging of the SGI1-B-Ec1 structure to the chromosomal context by sequencing reads.

samtools index AVC96_SGI1-B-EC1.bam

Long reads were mapped to the same SGI1-B-Ec1 sequence using minimap2 to check for the depth across the structure. This was done using the following commands:

minimap2 -ax map-pb SGI1-B-EC1.fa AVC96_filtered.fastq.gz > SGI1-B-EC1.sam

samtools view -hF 0x904 SGI1-B-EC1.sam | samtools sort -@8 -T SGI1-B-EC1.fa - -o SGI1-B-EC1.sam

samtools depth SGI1-B-EC1.sam > SGI1-B-EC1_long-read_depth.txt

Again, Tablet was used to confirm that PacBio reads bridged the junctions between the SGI1-B-Ec1 element and the chromosome of AVC96, however due to a difference in minimap2 output format the alignment file had to be indexed using tabix after zipping the file with bgzip.

bgzip SGI1-B-EC1.sam; tabix -p sam SGI1-B-EC1.sam.gz

Snapgene and SGI1-B-Ec1 schematic annotation

Snapgene annotations for SGI1-B were created using a GenBank file as a template before custom colouring and labelling of ORFs. Snapgene annotations for SGI1-B-Ec1 were created through the importation of annotated features from a GenBank file of SGI1B before subsequent confirmation using Megablast with SGI1B (Accession No.: KU987430) as a reference sequence. ISEc43, the insertion sequence inserted into S023, was identified using ISfinder (https://isfinder.biotoul.fr/). These annotations were used to make schematics of SGI1-B-Ec1 and SGI1-B which were added atop and below the EasyFig bitmap image, respectively, using image editing software.

Program sources and versions

Program	Version	URL	Manuscript Reference
Minimap2	2.12-r827	https://github.com/lh3/minimap2	(248)
BWA	0.7.17-r1188	https://github.com/lh3/bwa	(155)
Samtools	1.9	http://samtools.sourceforge.net/	(156)
EasyFig	2.2.3	http://mjsull.github.io/Easyfig/	(249)
Tablet	1.17.08.17	https://ics.hutton.ac.uk/tablet/	(250)
Unicycler	V0.3.0b	https://github.com/rrwick/unicycler	(251)
Filtlong	v0.2.0	https://github.com/rrwick/Filtlong	N/A
Snapgene	4.1.9	http://www.snapgene.com/	N/A

Chapter 8: References

1. Organization WH. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis. World Health Organization; 2017.

2. Roberts RR, Hota B, Ahmad I, Scott RD, 2nd, Foster SD, Abbasi F, et al. Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. Clin Infect Dis. 2009;49(8):1175-84.

3. Thorpe KE, Joski P, Johnston KJ. Antibiotic-resistant infection treatment costs have doubled since 2002, now exceeding $2 billion annually. Health Affairs. 2018;37(4):662-9.

4. Peirano G, van der Bij AK, Gregson DB, Pitout JD. Molecular epidemiology over an 11-year period (2000 to 2010) of extended-spectrum beta-lactamase-producing *Escherichia coli* causing bacteremia in a centralized Canadian region. J Clin Microbiol. 2012;50(2):294-9.

5. Reuland EA, Overdevest IT, Al Naiemi N, Kalpoe JS, Rijnsburger MC, Raadsen SA, et al. High prevalence of ESBL-producing Enterobacteriaceae carriage in Dutch community patients with gastrointestinal complaints. Clin Microbiol Infect. 2013;19(6):542-9.

6. Wang Y, Tian G-B, Zhang R, Shen Y, Tyrrell JM, Huang X, et al. Prevalence, risk factors, outcomes, and molecular epidemiology of *mcr-1*-positive Enterobacteriaceae in patients and healthy adults from China: an epidemiological and clinical study. The Lancet Infectious Diseases. 2017;17(4):390-9.

7. O'Neil J. Tackling drug-resistant infections globally: final report and recommendations: Review on Antimicrobial Resistance; 2016.

8. Bello-López JM, Cabrero-Martínez OA, Ibáñez-Cervantes G, Hernández-Cortez C, Pelcastre-Rodríguez LI, Gonzalez-Avila LU, et al. Horizontal gene transfer and its association with antibiotic resistance in the genus *Aeromonas* spp. Microorganisms. 2019;7(9):363.

9. Fillol-Salom A, Alsaadi A, de Sousa JAM, Zhong L, Foster KR, Rocha EP, et al. Bacteriophages benefit from generalized transduction. PLoS pathogens. 2019;15(7):e1007888.

10. Sun D. Pull in and push out: mechanisms of horizontal gene transfer in bacteria. Frontiers in microbiology. 2018;9:2154.

11. Weber RE, Pietsch M, Frühauf A, Pfeifer Y, Martin M, Luft D, et al. IS26-mediated transfer of blaNDM-1 as the main route of resistance transmission during a polyclonal, multispecies outbreak in a German hospital. Frontiers in Microbiology. 2019;10:2817.

12. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature reviews microbiology. 2005;3(9):711.

13. Cullik A, Pfeifer Y, Prager R, von Baum H, Witte W. A novel IS26 structure surrounds blaCTX-M genes in different plasmids from German clinical *Escherichia coli* isolates. J Med Microbiol. 2010;59(Pt 5):580-7.

14. Lee H, Doak TG, Popodi E, Foster PL, Tang H. Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. Nucleic acids research. 2016;44(15):7109-19.

15. Harmer CJ, Hall RM. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. mSphere. 2016;1(2).

16. Jain A, Srivastava P. Broad host range plasmids. FEMS microbiology letters. 2013;348(2):87-96.

17. Hayes F. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science. 2003;301(5639):1496-9.

18. Couturier M, Bex F, Bergquist PL, Maas WK. Identification and classification of bacterial plasmids. Microbiological reviews. 1988;52(3):375.

19. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. Journal of Antimicrobial Chemotherapy. 2018;73(5):1121-1137.

20. Mutai WC, Waiyaki PG, Kariuki S, Muigai AW. Plasmid profiling and incompatibility grouping of multidrug resistant *Salmonella enterica* serovar Typhi isolates in Nairobi, Kenya. BMC research notes. 2019;12(1):422.

21. Fernandes MR, McCulloch JA, Vianello MA, Moura Q, Perez-Chaparro PJ, Esposito F, et al. First Report of the Globally Disseminated IncX4 Plasmid Carrying the *mcr-1* Gene in a Colistin-Resistant *Escherichia coli* Sequence Type 101 Isolate from a Human Infection in Brazil. Antimicrob Agents Chemother. 2016;60(10):6415-7.

22. 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):2245-8.

23. Dolejska M, Masarikova M, Dobiasova H, Jamborova I, Karpiskova R, Havlicek M, et al. High prevalence of *Salmonella* and IMP-4-producing Enterobacteriaceae in the silver gull on Five Islands, Australia. J Antimicrob Chemother. 2016;71(1):63-70.

24. Moran RA, Hall RM. Evolution of regions containing antibiotic resistance genes in FII-2-FIB-1 ColV-Colla virulence plasmids. Microbial Drug Resistance. 2017.

25. Johnson TJ, Nolan LK. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol Mol Biol Rev. 2009;73(4):750-74.

26. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. Microbiol Mol Biol Rev. 1999;63(3):507-22.

27. Gillings M, Boucher Y, Labbate M, Holmes A, Krishnan S, Holley M, et al. The evolution of class 1 integrons and the rise of antibiotic resistance. J Bacteriol. 2008;190(14):5095-100.

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

29. Babakhani S, Oloomi M. Transposons: the agents of antibiotic resistance in bacteria. Journal of basic microbiology. 2018;58(11):905-17.

30. Snesrud E, McGann P, Chandler M. The birth and demise of the ISApl1-*mcr-1*-ISApl1 composite transposon: the vehicle for transferable colistin resistance. MBio. 2018;9(1):e02381-17.

31. Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM, Zhu YG. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. ISME J. 2015;9(6):1269-79.

32. Gillings MR, Stokes H. Are humans increasing bacterial evolvability? Trends in ecology & evolution. 2012;27(6):346-52.

33. Guédon G, Libante V, Coluzzi C, Payot S, Leblond-Bourget N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. Genes. 2017;8(11):337.

34. Doublet B, Boyd D, Mulvey MR, Cloeckaert A. The *Salmonella* genomic island 1 is an integrative mobilizable element. Mol Microbiol. 2005;55(6):1911-24.

35. Threlfall EJ. Epidemic *Salmonella* typhimurium DT 104—a truly international multiresistant clone. Journal of Antimicrobial Chemotherapy. 2000;46(1):7-10.

36. Hall RM. *Salmonella* genomic islands and antibiotic resistance in *Salmonella enterica*. Future microbiology. 2010;5(10):1525-38.

37. 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):4401-9.

38. Hamidian M, Holt KE, Hall RM. Genomic resistance island AGI1 carrying a complex class 1 integron in a multiply antibiotic-resistant ST25 *Acinetobacter baumannii* isolate. Journal of Antimicrobial Chemotherapy. 2015;70(9):2519-23.

39. Soliman AM, Shimamoto T, Nariya H, Shimamoto T. Emergence of *Salmonella* genomic island 1 variant SGI1-W in a clinical isolate of *Providencia stuartii* from Egypt. Antimicrobial agents and chemotherapy. 2018:AAC. 01793-18.

40. Doublet B, Golding GR, Mulvey MR, Cloeckaert A. Potential integration sites of the *Salmonella* genomic island 1 in *Proteus mirabilis* and other bacteria. Journal of antimicrobial chemotherapy. 2007;59(4):801-3.

41. Siebor E, de Curraize C, Amoureux L, Neuwirth C. Mobilization of the *Salmonella* genomic island SGI1 and the *Proteus* genomic island PGI1 by the A/C2 plasmid carrying bla TEM-24 harboured by various clinical species of Enterobacteriaceae. Journal of Antimicrobial Chemotherapy. 2016;71(8):2167-70.

42. de Curraize C, Neuwirth C, Bador J, Chapuis A, Amoureux L, Siebor E. Two new *Salmonella* genomic islands 1 from *Proteus mirabilis* and description of blaCTX-M-15 on a variant (SGI1-K7). Journal of Antimicrobial Chemotherapy. 2018.

43. Siebor E, Neuwirth C. The new variant of *Salmonella* genomic island 1 (SGI1-V) from a *Proteus mirabilis* French clinical isolate harbours blaVEB-6 and qnrA1 in the multiple antibiotic resistance region. J Antimicrob Chemother. 2011;66(11):2513-20.

44. Girlich D, Dortet L, Poirel L, Nordmann P. Integration of the blaNDM-1 carbapenemase gene into *Proteus* genomic island 1 (PGI1-PmPEL) in a *Proteus mirabilis* clinical isolate. J Antimicrob Chemother. 2015;70(1):98-102.

45. Wang X-C, Lei C-W, Kang Z-Z, Zhang Y, Wang H-N. IS26-mediated genetic rearrangements in *Salmonella* genomic island 1 of *Proteus mirabilis*. Frontiers in microbiology. 2019;10:2245.

46. Huguet KT, Gonnet M, Doublet B, Cloeckaert A. A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable *Salmonella* genomic island 1. Scientific reports. 2016;6:32285.

47. Donnenberg M. *Escherichia coli*: pathotypes and principles of pathogenesis: Academic Press; 2013.

48. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evolutionary applications. 2015;8(3):273-83.

49. Thanner S, Drissner D, Walsh F. Antimicrobial Resistance in Agriculture. MBio. 2016;7(2):e02227-15.

50. Woolhouse M, Ward M, van Bunnik B, Farrar J. Antimicrobial resistance in humans, livestock and the wider environment. Philos Trans R Soc Lond B Biol Sci. 2015;370(1670):20140083.

51. Reid CJ, Blau K, Jechalke S, Smalla K, Djordjevic SP. Whole Genome Sequencing of *Escherichia coli* From Store-Bought Produce. Frontiers in Microbiology. 2020;10:3050.

52. Control CfD. Outbreak of *E. coli* Infections Linked to Romaine Lettuce (2019). https://www.cdc.gov/ecoli/2019/o157h7-11-19/index.html: Centre for Disease Control; 2019.

53. Bouki C, Venieri D, Diamadopoulos E. Detection and fate of antibiotic resistant bacteria in wastewater treatment plants: a review. Ecotoxicology and environmental safety. 2013;91:1-9.

54. Fouz N, Pangesti KN, Yasir M, Al-Malki AL, Azhar EI, Hill-Cawthorne GA, et al. The Contribution of Wastewater to the Transmission of Antimicrobial Resistance in the Environment: Implications of Mass Gathering Settings. Tropical Medicine and Infectious Disease. 2020;5(1):33.

55. Chen PA, Hung CH, Huang PC, Chen JR, Huang IF, Chen WL, et al. Characteristics of CTX-M Extended-Spectrum beta-Lactamase-Producing *Escherichia coli* Strains Isolated from Multiple Rivers in Southern Taiwan. Appl Environ Microbiol. 2016;82(6):1889-97.

56. Harada K. [Antibiotic Residue in Environmental Water in Vietnam]. Yakugaku Zasshi. 2018;138(3):271-5.

57. Wu B, Qi Q, Zhang X, Cai Y, Yu G, Lv J, et al. Dissemination of *Escherichia coli* carrying plasmid-mediated quinolone resistance (PMQR) genes from swine farms to surroundings. Sci Total Environ. 2019;665:33-40.

58. Van Elsas JD, Semenov AV, Costa R, Trevors JT. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. The ISME journal. 2011;5(2):173-83.

59. Cytryn E. The soil resistome: the anthropogenic, the native, and the unknown. Soil Biology and Biochemistry. 2013;63:18-23.

60. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in molecular serotyping and subtyping of *Escherichia coli*. Frontiers in microbiology. 2016;7:644.

61. Buckles EL, Wang X, Lane MC, Lockatell CV, Johnson DE, Rasko DA, et al. Role of the K2 capsule in *Escherichia coli* urinary tract infection and serum resistance. The Journal of infectious diseases. 2009;199(11):1689-97.

62. Djordjevic SP, Hornitzky MA, Bailey G, Gill P, Vanselow B, Walker K, et al. Virulence properties and serotypes of Shiga toxin-producing *Escherichia coli* from healthy Australian slaughter-age sheep. J Clin Microbiol. 2001;39(5):2017-21.

63. Hornitzky MA, Mercieca K, Bettelheim KA, Djordjevic SP. Bovine feces from animals with gastrointestinal infections are a source of serologically diverse atypical enteropathogenic *Escherichia coli* and Shiga toxin-producing *E. coli* strains that commonly possess intimin. Appl Environ Microbiol. 2005;71(7):3405-12.

64. Bettelheim KA. The non-O157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit Rev Microbiol. 2007;33(1):67-87.

65. Nelson K, Selander RK. Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (gnd) in enteric bacteria. Proceedings of the National Academy of Sciences. 1994;91(21):10227-31.

66. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol. 2000;66(10):4555-8.

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

68. Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. J Infect. 2015;71(6):615-26.

69. Suelter CS, Hanson ND. OmpC regulation differs between ST131 and non-ST131 *Escherichia coli* clinical isolates and involves differential expression of the small RNA MicC. Journal of Antimicrobial Chemotherapy. 2020.

70. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, Gilbert GL. *Escherichia coli* sequence type 131 as a prominent cause of antibiotic resistance among urinary *Escherichia coli* isolates from reproductive-age women. Journal of clinical microbiology. 2013;51(10):3270-6.

71. Nicolas-Chanoine MH, Bertrand X, Madec JY. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev. 2014;27(3):543-74.

72. Cordoni G, Woodward MJ, Wu H, Alanazi M, Wallis T, La Ragione RM. Comparative genomics of European avian pathogenic E. Coli (APEC). BMC Genomics. 2016;17(1):960.

73. Manges AR, Harel J, Masson L, Edens TJ, Portt A, Reid-Smith RJ, et al. Multilocus sequence typing and virulence gene profiles associated with *Escherichia coli* from human and animal sources. Foodborne Pathog Dis. 2015;12(4):302-10.

74. Maluta RP, Logue CM, Casas MR, Meng T, Guastalli EA, Rojas TC, et al. Overlapped sequence types (STs) and serogroups of avian pathogenic (APEC) and human extra-intestinal pathogenic (ExPEC) *Escherichia coli* isolated in Brazil. PLoS One. 2014;9(8):e105016.

75. Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Group tAS, Achtman M. The user's guide to comparative genomics with EnteroBase, including case studies on transmissions of micro-clades of *Salmonella*, the phylogeny of ancient and modern Yersinia pestis genomes, and the core genomic diversity of all Escherichia. bioRxiv (preprint). 2019.

76. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS genetics. 2009;5(1).

77. Kohler CD, Dobrindt U. What defines extraintestinal pathogenic *Escherichia coli*? Int J Med Microbiol. 2011;301(8):642-7.

78. Johnson TJ, Wannemuehler Y, Doetkott C, Johnson SJ, Rosenberger SC, Nolan LK. Identification of minimal predictors of avian pathogenic *Escherichia coli* virulence for use as a rapid diagnostic tool. J Clin Microbiol. 2008;46(12):3987-96.

79. Fairbrother JM, Nadeau É, Gyles CL. *Escherichia coli* in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. Animal health research reviews. 2005;6(1):17-39.

80. 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.

81. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Dis Mon. 2003;49(2):53-70.

82. Akram M, Shahid M, Khan AU. Etiology and antibiotic resistance patterns of community-acquired urinary tract infections in JNMC Hospital Aligarh, India. Annals of clinical microbiology and antimicrobials. 2007;6(1):4.

83. Manges A, Natarajan P, Solberg O, Dietrich P, Riley L. The changing prevalence of drug-resistant *Escherichia coli* clonal groups in a community: evidence for community outbreaks of urinary tract infections. Epidemiology & Infection. 2006;134(2):425-31.

84. Cullen IM, Manecksha RP, McCullagh E, Ahmad S, O'Kelly F, Flynn RJ, et al. The changing pattern of antimicrobial resistance within 42 033 *Escherichia coli* isolates from nosocomial, community and urology patient-specific urinary tract infections, Dublin, 1999–2009. BJU international. 2012;109(8):1198-206.

85. Gyawali B, Ramakrishna K, Dhamoon AS. Sepsis: The evolution in definition, pathophysiology, and management. SAGE open medicine. 2019;7:2050312119835043.

86. Johnson TJ, Wannemuehler YM, Nolan LK. Evolution of the *iss* gene in *Escherichia coli*. Appl Environ Microbiol. 2008;74(8):2360-9.

87. Hejair HM, Ma J, Zhu Y, Sun M, Dong W, Zhang Y, et al. Role of outer membrane protein T in pathogenicity of avian pathogenic *Escherichia coli*. Res Vet Sci. 2017;115:109-16.

88. Usein CR, Papagheorghe R, Oprea M, Condei M, Straut M. Molecular characterization of bacteremic *Escherichia coli* isolates in Romania. Folia Microbiol (Praha). 2016;61(3):221-6.

89. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. Clin Microbiol Rev. 2019;32(3).

90. Wang S, Niu C, Shi Z, Xia Y, Yaqoob M, Dai J, et al. Effects of *ibeA* deletion on virulence and biofilm formation of avian pathogenic *Escherichia coli*. Infect Immun. 2011;79(1):279-87.

91. Ewers C, Li G, Wilking H, Kiessling S, Alt K, Antao EM, et al. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? Int J Med Microbiol. 2007;297(3):163-76.

92. Bingen E, Picard B, Brahimi N, Mathy S, Desjardins P, Elion J, et al. Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. J Infect Dis. 1998;177(3):642-50.

93. Dho-Moulin M, Fairbrother JM. Avian pathogenic *Escherichia coli* (APEC). Veterinary research. 1999;30(2-3):299-316.

94. Kabir S. Avian colibacillosis and salmonellosis: a closer look at epidemiology, pathogenesis, diagnosis, control and public health concerns. International journal of environmental research and public health. 2010;7(1):89-114.

95. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Nolan LK. Characterizing the APEC pathotype. Veterinary research. 2005;36(2):241-56.

96. Johnson TJ, Logue CM, Johnson JR, Kuskowski MA, Sherwood JS, Barnes HJ, et al. Associations between multidrug resistance, plasmid content, and virulence potential among extraintestinal pathogenic and commensal *Escherichia coli* from humans and poultry. Foodborne Pathog Dis. 2012;9(1):37-46.

97. Mellata M. Human and avian extraintestinal pathogenic *Escherichia coli*: infections, zoonotic risks, and antibiotic resistance trends. Foodborne pathogens and disease. 2013;10(11):916-32.

98. Manges AR, Smith SP, Lau BJ, Nuval CJ, Eisenberg JN, Dietrich PS, et al. Retail meat consumption and the acquisition of antimicrobial resistant *Escherichia coli* causing urinary tract infections: a case–control study. Foodborne pathogens and disease. 2007;4(4):419-31.

99. Mora A, Lopez C, Herrera A, Viso S, Mamani R, Dhabi G, 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):347-52.

100. Wu G, Ehricht R, Mafura M, Stokes M, Smith N, Pritchard GC, et al. *Escherichia coli* isolates from extraintestinal organs of livestock animals harbour diverse virulence genes and belong to multiple genetic lineages. Vet Microbiol. 2012;160(1-2):197-206.

101. McKinnon J, Chowdhury PR, Djordjevic SP. Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis. International journal of antimicrobial agents. 2018;52(3):430-5.

102. Johnson TJ, Jordan D, Kariyawasam S, Stell AL, Bell NP, Wannemuehler YM, et al. Sequence analysis and characterization of a transferable hybrid plasmid encoding multidrug resistance and enabling zoonotic potential for extraintestinal *Escherichia coli*. Infect Immun. 2010;78(5):1931-42.

103. Skyberg JA, Johnson TJ, Johnson JR, Clabots C, Logue CM, Nolan LK. 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):6287-92.

104. Johnson TJ, Thorsness JL, Anderson CP, Lynne AM, Foley SL, Han J, et al. Horizontal gene transfer of a ColV plasmid has resulted in a dominant avian clonal type of *Salmonella enterica* serovar Kentucky. PLoS One. 2010;5(12):e15524.

105. Alexandratos N, Bruinsma J. World agriculture towards 2030/2050: the 2012 revision. 2012.

106. Bradley P, Den Bakker H, Rocha E, McVean G, Iqbal Z. Real-time search of all bacterial and viral genomic data. 2019.

107. Downing T. Tackling Drug Resistant Infection Outbreaks of Global Pandemic *Escherichia coli* ST131 Using Evolutionary and Epidemiological Genomics. Microorganisms. 2015;3(2):236-67.

108. Rabsch W, Prager R, Koch J, Stark K, Roggentin P, Bockemühl J, et al. Molecular epidemiology of *Salmonella enterica* serovar Agona: characterization of a diffuse outbreak caused by aniseed-fennel-caraway infusion. Epidemiology & Infection. 2005;133(5):837-44.

109. Solà-Ginés M, Cameron-Veas K, Badiola I, Dolz R, Majó N, Dahbi G, et al. Diversity of multi-drug resistant avian pathogenic *Escherichia coli* (APEC) causing outbreaks of colibacillosis in broilers during 2012 in Spain. PloS one. 2015;10(11):e0143191.

110. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comput Biol. 2013;9(10):e1003285.

111. Tan TW, Tong JC, Khan AM, de Silva M, Lim KS, Ranganathan S. Advancing standards for bioinformatics activities: persistence, reproducibility, disambiguation and Minimum Information About a Bioinformatics investigation (MIABi). BMC genomics. 2010;11(S4):S27.

112. Warnke-Sommer J, Ali H. Graph mining for next generation sequencing: leveraging the assembly graph for biological insights. BMC genomics. 2016;17(1):340.

113. Tillett D, Neilan BA. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. Journal of Phycology. 2000;36(1):251-8.

114. Clermont O, Gordon D, Denamur E. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. Microbiology. 2015;161(Pt 5):980-8.

115. Tan C, Tang X, Zhang X, Ding Y, Zhao Z, Wu B, et al. Serotypes and virulence genes of extraintestinal pathogenic *Escherichia coli* isolates from diseased pigs in China. Vet J. 2012;192(3):483-8.

116. Jakobsen L, Kurbasic A, Skjøt-Rasmussen L, Ejrnæs K, Porsbo LJ, Pedersen K, et al. *Escherichia coli* isolates from broiler chicken meat, broiler chickens, pork, and pigs share phylogroups and antimicrobial resistance with community-dwelling humans and patients with urinary tract infection. Foodborne pathogens and disease. 2010;7(5):537-47.

117. Jakobsen L, Garneau P, Bruant G, Harel J, Olsen S, Porsbo L, et al. Is *Escherichia coli* urinary tract infection a zoonosis? Proof of direct link with production animals and meat. European journal of clinical microbiology & infectious diseases. 2012;31(6):1121-9.

118. Gupta K, Hooton TM, Stamm WE. Increasing antimicrobial resistance and the management of uncomplicated community-acquired urinary tract infections. Annals of internal medicine. 2001;135(1):41-50.

119. Tivendale KA, Noormohammadi AH, Allen JL, Browning GF. The conserved portion of the putative virulence region contributes to virulence of avian pathogenic *Escherichia coli*. Microbiology. 2009;155(Pt 2):450-60.

120. Tivendale KA, Allen JL, Ginns CA, Crabb BS, Browning GF. Association of *iss* and *iucA,* but not tsh, with plasmid-mediated virulence of avian pathogenic *Escherichia coli*. Infection and immunity. 2004;72(11):6554-60.

121. Delicato ER, de Brito BG, Gaziri LCJ, Vidotto MC. Virulence-associated genes in *Escherichia coli* isolates from poultry with colibacillosis. Veterinary microbiology. 2003;94(2):97-103.

122. 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. 2006;188(2):745-58.

123. Pfaff-McDonough SJ, Horne SM, Giddings CW, Ebert JO, Doetkott C, Smith MH, et al. Complement resistance-related traits among *Escherichia coli* isolates from apparently healthy birds and birds with colibacillosis. Avian diseases. 2000;44(1):23-33.

124. Zhao L, Gao S, Huan H, Xu X, Zhu X, Yang W, et al. Comparison of virulence factors and expression of specific genes between uropathogenic *Escherichia coli* and avian pathogenic *E. coli* in a murine urinary tract infection model and a chicken challenge model. Microbiology. 2009;155(5):1634-44.

125. Stromberg ZR, Johnson JR, Fairbrother JM, Kilbourne J, Van Goor A, Curtiss 3rd R, et al. Evaluation of *Escherichia coli* isolates from healthy chickens to determine their potential risk to poultry and human health. PloS one. 2017;12(7):e0180599.

126. Tivendale KA, Logue CM, Kariyawasam S, Jordan D, Hussein A, Li G, et al. Avian-pathogenic *Escherichia coli* strains are similar to neonatal meningitis *E. coli* strains and are able to cause meningitis in the rat model of human disease. Infect Immun. 2010;78(8):3412-9.

127. Chanteloup NK, Porcheron G, Delaleu B, Germon P, Schouler C, Moulin-Schouleur M, et al. The extra-intestinal avian pathogenic *Escherichia coli* strain BEN2908 invades avian and human epithelial cells and survives intracellularly. Vet Microbiol. 2011;147(3-4):435-9.

128. Johnson JR, Delavari P, Kuskowski M, Stell AL. Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. The Journal of infectious diseases. 2001;183(1):78-88.

129. Johnson JR, Murray AC, Gajewski A, Sullivan M, Snippes P, Kuskowski MA, 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):2161-8.

130. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Fakhr MK, Nolan LK. Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. Microbiology. 2005;151(6):2097-110.

131. Manges AR. *Escherichia coli* and urinary tract infections: the role of poultry-meat. Clin Microbiol Infect. 2016;22(2):122-9.

132. Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L, Galanakis C, et al. Food reservoir for *Escherichia coli* causing urinary tract infections. Emerging infectious diseases. 2010;16(1):88.

133. Manges AR, Johnson JR. Food-borne origins of *Escherichia coli* causing extraintestinal infections. Clin Infect Dis. 2012;55(5):712-9.

134. Messenger AM, Barnes AN, Gray GC. Reverse zoonotic disease transmission (zooanthroponosis): a systematic review of seldom-documented human biological threats to animals. PloS one. 2014;9(2):e89055.

135. David E. Swayne LKN, H. John Barnes, Jean-Pierre Vaillancourt,Tahseen Abdul-Aziz, and Catherine M. Logue. Diseases of Poultry, 13th Edition: Chapter 18 - Colibacillosis. 13 ed: Wiley-Blackwell; 2013. 751-806 p.

136. Guabiraba R, Schouler C. Avian colibacillosis: still many black holes. FEMS Microbiol Lett. 2015;362(15):fnv118.

137. Boyd EF, Hartl DL. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. J Bacteriol. 1998;180(5):1159-65.

138. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Molecular biology and evolution. 2007;24(11):2373-84.

139. 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. Frontiers in microbiology. 2013;4:86.

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

141. Márquez C, Labbate M, Raymondo C, Fernández J, Gestal AM, Holley M, et al. Urinary tract infections in a South American population: dynamic spread of class 1 integrons and multidrug resistance by homologous and site-specific recombination. Journal of clinical microbiology. 2008;46(10):3417-25.

142. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics. 2015;31(4):587-9.

143. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic acids research. 2006;34(suppl_1):D32-D6.

144. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic acids research. 2005;33(suppl_1):D325-D8.

145. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome medicine. 2014;6(11):90.

146. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microbial Genomics. 2017;3(10).

147. Bagel S, Hüllen V, Wiedemann B, Heisig P. Impact of *gyrA* and *parC M*utations on Quinolone Resistance, Doubling Time, and Supercoiling Degree of *Escherichia coli*. Antimicrobial agents and chemotherapy. 1999;43(4):868-75.

148. Seemann T. Snippy: fast bacterial variant calling from NGS reads. 2015.

149. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276-8.

150. Darling AE, Jospin G, Lowe E, Matsen IV FA, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ. 2014;2:e243.

151. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one. 2010;5(3):e9490.

152. Wyrsch E, Chowdhury PR, Abraham S, Santos J, Darling AE, Charles IG, et al. Comparative genomic analysis of a multiple antimicrobial resistant enterotoxigenic *E. coli* O157 lineage from Australian pigs. BMC genomics. 2015;16(1):165.

153. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2006;23(1):127-8.

154. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8(1):28-36.

155. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

156. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

157. Pires-dos-Santos T, Bisgaard M, Christensen H. Genetic diversity and virulence profiles of *Escherichia coli* causing salpingitis and peritonitis in broiler breeders. Vet Microbiol. 2013;162(2-4):873-80.

158. Fernandes MR, Sellera FP, Moura Q, Souza TA, Lincopan N. Draft genome sequence of a CTX-M-8, CTX-M-55 and FosA3 co-producing *Escherichia coli* ST117/B2 isolated from an asymptomatic carrier. Journal of global antimicrobial resistance. 2018;12:183-4.

159. Braga JFV, Chanteloup NK, Trotereau A, Baucheron S, Guabiraba R, Ecco R, et al. Diversity of *Escherichia coli* strains involved in vertebral osteomyelitis and arthritis in broilers in Brazil. BMC Vet Res. 2016;12(1):140.

160. Kemmett K, Williams NJ, Chaloner G, Humphrey S, Wigley P, Humphrey T. The contribution of systemic *Escherichia coli* infection to the early mortalities of commercial broiler chickens. Avian Pathol. 2014;43(1):37-42.

161. Ronco T, Stegger M, Olsen RH, Sekse C, Nordstoga AB, Pohjanvirta T, et al. Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. BMC Genomics. 2017;18(1):13.

162. Olsen RH, Chadfield MS, Christensen JP, Scheutz F, Christensen H, Bisgaard M. Clonality and virulence traits of *Escherichia coli* associated with haemorrhagic septicaemia in turkeys. Avian Pathol. 2011;40(6):587-95.

163. Knobl T, Moreno AM, Paixao R, Gomes TA, Vieira MA, da Silva Leite D, et al. Prevalence of avian pathogenic *Escherichia coli* (APEC) clone harboring *sfa* gene in Brazil. ScientificWorldJournal. 2012;2012:437342.

164. Muller A, Stephan R, Nuesch-Inderbinen M. Distribution of virulence factors in ESBL-producing *Escherichia coli* isolated from the environment, livestock, food and humans. Sci Total Environ. 2016;541:667-72.

165. Trobos M, Christensen H, Sunde M, Nordentoft S, Agerso Y, Simonsen GS, et al. Characterization of sulphonamide-resistant *Escherichia coli* using comparison of sul2 gene sequences and multilocus sequence typing. Microbiology. 2009;155(Pt 3):831-6.

166. Alikhan N-F, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of *Salmonella*. PLoS genetics. 2018;14(4):e1007261.

167. Matter LB, Spricigo DA, Tasca C, Vargas AC. Invasin *gimB* found in a bovine intestinal *Escherichia coli* with an adherent and invasive profile. Braz J Microbiol. 2015;46(3):875-8.

168. Hertz FB, Nielsen JB, Schonning K, Littauer P, Knudsen JD, Lobner-Olesen A, et al. "Population structure of drug-susceptible,-resistant and ESBL-producing *Escherichia coli* from community-acquired urinary tract". BMC Microbiol. 2016;16:63.

169. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. Clin Microbiol Infect. 2014;20(5):380-90.

170. Stephens CM, Adams-Sapper S, Sekhon M, Johnson JR, Riley LW. Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic *Escherichia coli* Sequence Type 95 Strains. mSphere. 2017;2(2).

171. Abraham S, Jordan D, Wong HS, Johnson JR, Toleman MA, Wakeham DL, 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):273-7.

172. Guo S, Wakeham D, Brouwers HJ, Cobbold RN, Abraham S, Mollinger JL, et al. Human-associated fluoroquinolone-resistant *Escherichia coli* clonal lineages, including ST354, isolated from canine feces and extraintestinal infections in Australia. Microbes Infect. 2015;17(4):266-74.

173. Awawdeh L. Studies on avian pathogenic *Escherichia coli* in commercial broiler Chicken in South East Queensland. The University of Queensland: The University of Queensland; 2018.

174. Younis G, Awad A, Mohamed N. Phenotypic and genotypic characterization of antimicrobial susceptibility of avian pathogenic *Escherichia coli* isolated from broiler chickens. Vet World. 2017;10(10):1167-72.

175. Liao XP, Xia J, Yang L, Li L, Sun J, Liu YH, et al. Characterization of CTX-M-14-producing *Escherichia coli* from food-producing animals. Front Microbiol. 2015;6:1136.

176. Lima Barbieri N, Nielsen DW, Wannemuehler Y, Cavender T, Hussein A, Yan SG, et al. *mcr-1* identified in Avian Pathogenic *Escherichia coli* (APEC). PLoS One. 2017;12(3):e0172997.

177. Perreten V, Strauss C, Collaud A, Gerber D. Colistin resistance gene *mcr-1* in avian-pathogenic *Escherichia coli* in South Africa. Antimicrobial agents and chemotherapy. 2016;60(7):4414-5.

178. Trung NV, Matamoros S, Carrique-Mas JJ, Nghia NH, Nhung NT, Chieu TTB, et al. Zoonotic transmission of *mcr-1* colistin resistance gene from small-scale poultry farms, Vietnam. Emerging infectious diseases. 2017;23(3):529.

179. Wang Y, Zhang R, Li J, Wu Z, Yin W, Schwarz S, et al. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. Nat Microbiol. 2017;2:16260.

180. Authority AAPaVM. Antibiotic resistance in animals. 2017.

181. Ho PL, Lo WU, Yeung MK, Lin CH, Chow KH, Ang I, et al. Complete sequencing of pNDM-HK encoding NDM-1 carbapenemase from a multidrug-resistant *Escherichia coli* strain isolated in Hong Kong. PloS one. 2011;6(3):e17989.

182. Mangat CS, Bekal S, Irwin RJ, Mulvey MR. A novel hybrid plasmid carrying multiple antimicrobial resistance and virulence genes in *Salmonella enterica* serovar Dublin. Antimicrobial agents and chemotherapy. 2017;61(6):e02601-16.

183. Porse A, Schønning 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):2860-73.

184. Harmer CJ, Hall RM. IS26-Mediated Precise Excision of the IS26-*aphA1a* Translocatable Unit. MBio. 2015;6(6):e01866-15.

185. Harmer CJ, Hall RM. Targeted conservative formation of cointegrates between two DNA molecules containing IS26 occurs via strand exchange at either IS end. Molecular microbiology. 2017;106(3):409-18.

186. 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. Microb Drug Resist. 2010;16(3):197-202.

187. Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP, Walker MJ. 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.

188. Barbieri NL, de Oliveira AL, Tejkowski TM, Pavanelo DB, Matter LB, Pinheiro SR, et al. Molecular characterization and clonal relationships among *Escherichia coli* strains isolated from broiler chickens with colisepticemia. Foodborne Pathog Dis. 2015;12(1):74-83.

189. Cunha MP, de Oliveira MG, de Oliveira MC, da Silva KC, Gomes CR, Moreno AM, et al. Virulence profiles, phylogenetic background, and antibiotic resistance of *Escherichia coli* isolated from turkeys with airsacculitis. ScientificWorldJournal. 2014;2014:289024.

190. Dissanayake DR, Octavia S, Lan R. Population structure and virulence content of avian pathogenic *Escherichia coli* isolated from outbreaks in Sri Lanka. Vet Microbiol. 2014;168(2-4):403-12.

191. Xu WY, Li YJ, Fan C. Different loci and mRNA copy number of the increased serum survival gene of *Escherichia coli*. Canadian journal of microbiology. 2018;64(2):147-54.

192. Lynne AM, Skyberg JA, Logue CM, Doetkott C, Foley SL, Nolan LK. Characterization of a series of transconjugant mutants of an avian pathogenic *Escherichia coli* isolate for resistance to serum complement. Avian diseases. 2007;51(3):771-6.

193. Pramoonjago P, Kaneko M, Kinoshita T, Ohtsubo E, Takeda J, Hong K, et al. Role of TraT protein, an anticomplementary protein produced in *Escherichia coli* by R100 factor, in serum resistance. The Journal of Immunology. 1992;148(3):827-36.

194. Lynne AM, Kariyawasam S, Wannemuehler Y, Johnson TJ, Johnson SJ, Sinha AS, et al. Recombinant Iss as a potential vaccine for avian colibacillosis. Avian diseases. 2012;56(1):192-9.

195. Chen YM, Wright PJ, Lee CS, Browning GF. Uropathogenic virulence factors in isolates of *Escherichia coli* from clinical cases of canine pyometra and feces of healthy bitches. Vet Microbiol. 2003;94(1):57-69.

196. Pourbakhsh SA, Dho-Moulin M, Bree A, Desautels C, Martineau-Doize B, Fairbrother JM. Localization of the in vivo expression of P and F1 fimbriae in chickens experimentally inoculated with pathogenic *Escherichia coli*. Microbial pathogenesis. 1997;22(6):331-41.

197. Lund B, Lindberg F, Marklund BI, Normark S. Tip proteins of pili associated with pyelonephritis: new candidates for vaccine development. Vaccine. 1988;6(2):110-2.

198. Kariyawasam S, Johnson TJ, Nolan LK. The pap operon of avian pathogenic *Escherichia coli* strain O1:K1 is located on a novel pathogenicity island. Infect Immun. 2006;74(1):744-9.

199. Rice JC, Peng T, Spence JS, Wang HQ, Goldblum RM, Corthesy B, et al. Pyelonephritic *Escherichia coli* expressing P fimbriae decrease immune response of the mouse kidney. Journal of the American Society of Nephrology : JASN. 2005;16(12):3583-91.

200. Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, et al. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. Appl Environ Microbiol. 2007;73(6):1976-83.

201. Moran RA, Hall RM. Evolution of Regions Containing Antibiotic Resistance Genes in FII-2-FIB-1 ColV-Colla Virulence Plasmids. Microb Drug Resist. 2018;24(4):411-21.

202. Hammond DS, Harris T, Bell J, Turnidge J, Giffard PM. Selection of SHV extended-spectrum-β-lactamase-dependent cefotaxime and ceftazidime resistance in Klebsiella pneumoniae requires a plasmid-borne blaSHV gene. Antimicrobial agents and chemotherapy. 2008;52(2):441-5.

203. Schmieger H, Schicklmaier P. Transduction of multiple drug resistance of *Salmonella enterica* serovar typhimurium DT104. FEMS Microbiol Lett. 1999;170(1):251-6.

204. Carraro N, Matteau D, Luo P, Rodrigue S, Burrus V. The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. PLoS Genet. 2014;10(10):e1004714.

205. Kiss J, Nagy B, Olasz F. Stability, entrapment and variant formation of *Salmonella* genomic island 1. PloS one. 2012;7(2):e32497.

206. 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):282-9.

207. Siebor E, de Curraize C, Amoureux L, Neuwirth C. Mobilization of the *Salmonella* genomic island SGI1 and the *Proteus* genomic island PGI1 by the A/C2 plasmid carrying blaTEM-24 harboured by various clinical species of Enterobacteriaceae. J Antimicrob Chemother. 2016;71(8):2167-70.

208. Schultz E, Barraud O, Madec J-Y, Haenni M, Cloeckaert A, Ploy M-C, et al. Multidrug resistance *Salmonella* genomic island 1 in a *Morganella morganii* subsp. morganii human clinical isolate from France. MSphere. 2017;2(2):e00118-17.

209. Ebner P, Garner K, Mathew A. Class 1 integrons in various *Salmonella enterica* serovars isolated from animals and identification of genomic island SGI1 in *Salmonella enterica* var. Meleagridis. J Antimicrob Chemother. 2004;53(6):1004-9.

210. Levings RS, Partridge SR, Djordjevic SP, Hall RM. SGI1-K, a variant of the SGI1 genomic island carrying a mercury resistance region, in *Salmonella enterica* serovar Kentucky. Antimicrobial agents and chemotherapy. 2007;51(1):317-23.

211. Djordjevic SP, Cain AK, Evershed NJ, Falconer L, Levings RS, Lightfoot D, et al. Emergence and evolution of multiply antibiotic-resistant *Salmonella enterica* serovar Paratyphi B D-tartrate-utilizing strains containing SGI1. Antimicrob Agents Chemother. 2009;53(6):2319-26.

212. 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):1750-6.

213. Siebor E, de Curraize C, Neuwirth C. Identification of AGI1-A, a variant of *Acinetobacter* genomic island 1 (AGI1), in a French clinical isolate belonging to the *Enterobacter cloacae* complex. J Antimicrob Chemother. 2019;74(2):311-4.

214. Soliman AM, Shimamoto T, Nariya H, Shimamoto T. Emergence of *Salmonella* Genomic Island 1 Variant SGI1-W in a Clinical Isolate of *Providencia stuartii* from Egypt. Antimicrob Agents Chemother. 2019;63(1).

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

216. Bradley P, Den Bakker H, Rocha E, McVean G, Iqbal Z. Real-time search of all bacterial and viral genomic data. BioRxiv. 2017:234955.

217. Malcova M, Hradecka H, Karpiskova R, Rychlik I. Biofilm formation in field strains of *Salmonella enterica* serovar Typhimurium: identification of a new colony morphology type and the role of SGI1 in biofilm formation. Vet Microbiol. 2008;129(3-4):360-6.

218. Wang XC, Lei CW, Kang ZZ, Zhang Y, Wang H-N. IS26-mediated genetic rearrangements in *Salmonella* genomic island 1 of *Proteus mirabilis*. Frontiers in Microbiology. 2019;10:2245.

219. Siebor E, de Curraize C, Neuwirth C. Identification of AGI1-A, a variant of *Acinetobacter* genomic island 1 (AGI1), in a French clinical isolate belonging to the *Enterobacter cloacae* complex. Journal of Antimicrobial Chemotherapy. 2018;74(2):311-4.

220. Cummins ML, Chowdhury PR, Marenda MS, Browning GF, Djordjevic SP. *Salmonella* Genomic Island 1B Variant Found in a Sequence Type 117 Avian Pathogenic *Escherichia coli* Isolate. mSphere. 2019;4(3):e00169-19.

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

222. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science. 2017;3:e104.

223. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research. 2014;43(3):e15-e.

224. Cummins ML, Reid CJ, Chowdhury PR, Bushell RN, Esbert N, Tivendale KA, et al. Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene. Microbial genomics. 2019;5(2).

225. de Curraize C, Siebor E, Neuwirth C, Hall RM. SGI0, a relative of *Salmonella* genomic islands SGI1 and SGI2, lacking a class 1 integron, found in *Proteus mirabilis*. Plasmid. 2019:102453.

226. Ellington MJ, Heinz E, Wailan AM, Dorman MJ, De Goffau M, Cain AK, et al. Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. Genome biology. 2019;20(1):1-16.

227. Kluytmans-van den Bergh MF, Rossen JW, Bruijning-Verhagen PC, Bonten MJ, Friedrich AW, Vandenbroucke-Grauls CM, et al. Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing Enterobacteriaceae. Journal of clinical microbiology. 2016;54(12):2919-27.

228. Partridge SR, Brown HJ, Stokes HW, Hall RM. Transposons Tn1696 and Tn21and their integrons In4 and In2 have independent origins. Antimicrobial agents and chemotherapy. 2001;45(4):1263-70.

229. Partridge SR, Recchia GD, Stokes HW, Hall RM. Family of class 1 integrons related to In4 from Tn1696. Antimicrobial agents and chemotherapy. 2001;45(11):3014-20.

230. Qiu D, Wei H, Tu Q, Yang Y, Xie M, Chen J, et al. Combined genomics and experimental analyses of respiratory characteristics of *Shewanella putrefaciens* W3-18-1. Appl Environ Microbiol. 2013;79(17):5250-7.

231. Prim N, Rivera A, Rodríguez-Navarro J, Español M, Turbau M, Coll P, et al. Detection of *mcr-1* colistin resistance gene in polyclonal *Escherichia coli* isolates in Barcelona, Spain, 2012 to 2015. Eurosurveillance. 2016;21(13):30183.

232. Vila J, Saez-Lopez E, Johnson JR, Romling U, Dobrindt U, Canton R, et al. *Escherichia coli*: an old friend with new tidings. FEMS Microbiol Rev. 2016;40(4):437-63.

233. Walsh TR, Wu Y. China bans colistin as a feed additive for animals. The Lancet Infectious Diseases. 2016;16(10):1102.

234. Agency EM. Updated advice on the use of colistin products in animals within the European Union: development of resistance and possible impact on human and animal health. EMA London, United Kingdom; 2016.

235. Jørgensen SL, Stegger M, Kudirkiene E, Lilje B, Poulsen LL, Ronco T, et al. Diversity and population overlap between avian and human *Escherichia coli* belonging to sequence type 95. Msphere. 2019;4(1):e00333-18.

236. Johnson TJ, Wannemuehler Y, Johnson SJ, Stell AL, Doetkott C, Johnson JR, et al. Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. Appl Environ Microbiol. 2008;74(22):7043-50.

237. Ghunaim H, Abu-Madi MA, Kariyawasam S. Advances in vaccination against avian pathogenic *Escherichia coli* respiratory disease: potentials and limitations. Vet Microbiol. 2014;172(1-2):13-22.

238. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, Nordstrom L, et al. *Escherichia coli* ST131-H22 as a foodborne uropathogen. MBio. 2018;9(4):e00470-18.

239. Bogema DR, McKinnon J, Liu M, Hitchick N, Miller N, Venturini C, et al. Whole-genome analysis of extraintestinal *Escherichia coli* sequence type 73 from a single hospital over a 2 year period identified different circulating clonal groups. Microbial Genomics. 2020;6(1).

240. Blanco JE, Blanco M, Blanco J. [Enterotoxigenic, verotoxigenic, and necrotoxigenic *Escherichia coli* in food and clinical samples. Role of animals as reservoirs of strains pathogenic for humans]. Microbiologia (Madrid, Spain). 1995;11(1):97-110.

241. Mora A, Viso S, Lopez C, Alonso MP, Garcia-Garrote F, Dabhi G, et al. Poultry as reservoir for extraintestinal pathogenic *Escherichia coli* O45:K1:H7-B2-ST95 in humans. Vet Microbiol. 2013;167(3-4):506-12.

242. Xia X, Meng J, McDermott PF, Zhao S. *Escherichia coli* from retail meats carry genes associated with uropathogenic *Escherichia coli*, but are weakly invasive in human bladder cell culture. J Appl Microbiol. 2011;110(5):1166-76.

243. Goetting V, Lee K, Tell LA. Pharmacokinetics of veterinary drugs in laying hens and residues in eggs: a review of the literature. Journal of veterinary pharmacology and therapeutics. 2011;34(6):521-56.

244. Mangat CS, Bekal S, Avery BP, Côté G, Daignault D, Doualla-Bell F, et al. Genomic investigation of the emergence of invasive multidrug-resistant *Salmonella enterica* serovar Dublin in humans and animals in Canada. Antimicrobial agents and chemotherapy. 2019;63(6):e00108-19.

245. Armand-Lefèvre L, Andremont A, Ruppé E. Travel and acquisition of multidrug-resistant Enterobacteriaceae. Medecine et maladies infectieuses. 2018;48(7):431-41.

246. Kantele A, Kuenzli E, Dunn SJ, Dance DA, Newton PN, Davong V, et al. Real-time sampling of travelers shows intestinal colonization by multidrug-resistant bacteria to be a dynamic process with multiple transient acquisitions. bioRxiv. 2019:827915.

247. Kantele A, Lääveri T, Mero S, Häkkinen IM, Kirveskari J, Johnston BD, et al. Despite Predominance of Uropathogenic/Extraintestinal Pathotypes Among Travel-acquired Extended-spectrum β-Lactamase–producing *Escherichia coli*, the Most Commonly Associated Clinical Manifestation Is Travelers' Diarrhea. Clinical Infectious Diseases. 2020;70(2):210-8.

248. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;1:7.

249. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009-10.

250. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, et al. Tablet—next generation sequence assembly visualization. Bioinformatics. 2009;26(3):401-2.

251. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS computational biology. 2017;13(6):e1005595.