UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Genomic analysis of uropathogenic Escherichia coli from hospital patients in Australia

by Dmitriy Li

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Prof Steven P. Djordjevic and Dr Veronica M. Jarocki

University of Technology Sydney
Faculty of Science

July 2023

Certificate of Original Authorship

I, **Dmitriy Li** declare that this thesis is submitted in fulfilment of the requirements for the award of **Doctor of Philosophy**, in the **Australian Institute for Microbiology & Infection (AIMI)** 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: **23 06 2023**

Acknowledgments

First, I would like to thank my supervisors, Distinguished Prof Steven Djordjevic and Dr Veronica Jarocki, who made this project possible. Your expertise, guidance, support, and training were immensely appreciated and allowed me to finish. Also, I would also like to thank Dr Timothy Kudinha and Orange Base Hospital, for providing me with a wonderful collection of clinical *Escherichia coli* isolates that I had the privilege to work on during my candidature.

Thank you to all the members of Djordjevic lab, past and present. I would particularly like to thank Dr Cameron Reid and Dr Max Cummins for their invaluable help and guidance in the new to me field of bioinformatics, as well as for the scripts and databases provided to me during my candidature.

Enormous thanks to the University of Technology Sydney (UTS), the Australian Institute for Microbiology and Infection (AIMI), Ausgem and UTS eResearch as well, each of which has been incredible in providing financial and technical support throughout my project.

Finally, I would like to thank my family. Thank you to my parents for your financial and emotional support and for always reminding me how proud you are of me.

Statement of Thesis Format

This thesis is presented by compilation. Chapters 4, 5 and 6 are results chapters; Chapters 5 and 6 have been published in a peer-reviewed journal, chapter 4 is a manuscript in preparation for submission. Figure and Table numbers have been edited from original publications to match the chapter numbering. Minor changes to the published text were also made so abbreviations used are consistent throughout the document. Due to the thesis margin requirements, it is recommended to view the figures in the original publications for Chapters 5 and 6.

List of Publications

Published

Title: Genomic analysis of trimethoprim-resistant extraintestinal pathogenic *Escherichia coli* and recurrent urinary tract infections

Authors: **Dmitriy Li**[1], Cameron J. Reid[1], Timothy Kudinha[2,3], Veronica M. Jarocki[1,] and Steven P. Djordjevic[1]

Published in Microbial Genomics, November 2020, DOI: 10.1099/mgen.0.000475

Author affiliations:

[1] Ithree Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia
[2]NSW Health Pathology, Microbiology, Orange Hospital, Orange, NSW 2800, Australia
[3]School of Biomedical Sciences, Charles Sturt University, Orange, NSW 2800, Australia.

Title: Genomic comparisons of *Escherichia coli* ST131 from Australia

Authors: **Dmitriy Li**[1], Ethan R. Wyrsch[1], Paarthiphan Elankumaran[1], Monika Dolejska[2,3,4,5], Marc S. Marenda[6], Glenn F. Browning[6], Rhys N. Bushell[6], Jessica McKinnon[1], Piklu Roy Chowdhury[1], Nola Hitchick[7], Natalie Miller[7], Erica Donner[8], Barbara Drigo[8], Dave Baker[9], Ian G. Charles[9], Timothy Kudinha[10], Veronica M. Jarocki[1,], and Steven Philip Djordjevic[1]

Published in Microbial Genomics, December 2021, DOI: 10.1099/mgen.0.000721

Author affiliations:

[1] iThree Institute, University of Technology Sydney, Ultimo, NSW, Australia;
[2] CEITEC VETUNI, University of Veterinary Sciences Brno, Brno, Czech Republic;
[3] Department of Biology and Wildlife Disease, Faculty of Veterinary Hygiene and Ecology, University of Veterinary Sciences Brno, Czech Republic;
[4] Biomedical Center, Charles University, Czech Republic;
[5] Department of Clinical Microbiology and Immunology, Institute of Laboratory Medicine, The University Hospital Brno, Brno, Czech Republic;
[6] Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Victoria, Australia;
[7] San Pathology, Sydney Adventist Hospital, Wahroonga, NSW 2076, Australia;
[8] Future Industries Institute, University of South Australia, Adelaide, South Australia, Australia;
[9] Quadram Institute, Norwich, UK;
[10] Central West Pathology Laboratory, Charles Sturt University, Orange, NSW, 2800, Australia.

Title: Whole Genome Sequencing of Avian Pathogenic Escherichia coli Causing Bacterial Chondronecrosis and Osteomyelitis in Australian Poultry

Authors: Max L. Cummins[1,2], **Dmitriy Li**[1,2], Aeman Ahmad[1,2], Rhys Bushell[3], Amir H. Noormohammadi[3], Dinidu S. Wijesurendra[3], Andrew Stent[4], Marc S. Marenda[3] and Steven P. Djordjevic[1,2]

Published in Microorganisms, June 2023,DOI: 10.3390/microorganisms11061513

Author affiliations:

[1] Australian Institute for Microbiology and Infection, University of Technology Sydney, Ultimo, NSW 2007, Australia
[2] The Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[3] Faculty of Science, The University of Melbourne, Parkville, VIC 3010, Australia
[4] Gribbles Veterinary Pathology, Clayton, VIC 3168, Australia

Accepted for publication

Title: Dominance of *Escherichia coli* sequence types ST73, ST95, ST127, and ST131 in historic urine isolates: a genomic analysis of antimicrobial resistance and virulence linked to F plasmids

Authors: **Dmitriy Li**[1,2], Paarthiphan Elankumaran[1,2], Timothy Kudinha[3], Amanda Kidsley[4], Darren Trott[4], Veronica M. Jarocki[1,2,], Steven P. Djordjevic[1,2,]

Accepted for publication in Microbial Genomics, 21 June 2023

Author affiliations:

[1] Australian Institute for Microbiology & Infection, University of Technology Sydney, Ultimo, NSW, Australia
[2] Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, NSW, Australia
[3] Central West Pathology Laboratory, Charles Sturt University, Orange, NSW, Australia
[4] School of Animal and Veterinary Science, The University of Adelaide, Adelaide, South Australia, Australia

In Preparation

Title: Phylogenomic analysis of a global cohort of *Escherichia coli* ST38: evidence of interspecies and environmental transmission

Authors: Piklu Roy Chowdhury[1,2*], Priyanka Hastak[1,2], Ethan Wyrsch[1], **Dmitriy Li**[1], Paarthiphan Elankumaran[1], Monika Dolejska, Glen F. Browning, Mark Marenda, Thomas Gottlieb[3,4], Elaine Cheong[3,4], John Merlino[3,4], Garry S. A. Myers[1] and Steven P. Djordjevic[1,2*]

Author affiliations:

[1]The Ithree Institute, University of Technology Sydney, City Campus, Ultimo, NSW 2007, Australia.
[2] Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia
[3]Department of Microbiology and Infectious Diseases, Concord Hospital and NSW Health Pathology, Hospital Road, Concord 2139, NSW, Australia.
[4] Faculty of Medicine, University of Sydney, NSW Australia.

Title: First report of virulent and emerging sequence types of multidrug resistant *Escherichia coli* in hospitalized patients in a university health center

Authors: Wozniak Aniela[1], Brito Bárbara[2], García Patricia[1,3], Jarocki Veronica[2], **Li Dmitriy**[2], Djordjevic Steven[2]

Author affiliations:

[1]Department of Clinical Laboratories. Medicine School. Pontifical Catholic University of Chile
[2]The ithree Institute. Faculty of Science. University of Technology Sydney
[3]Millennium Initiative for Collaborative Research on Bacterial Resistance, MICROB-R, Santiago, Chile

Title: Genomic analysis of Australian wastewater *Escherichia coli*

Authors: Veronica M. Jarocki[1,2], **Dmitriy Li**[1,2], Daniel R. Bogema[2,3], Jerald Yam, Faisal Hai[4], Cheryl Jenkins[2,3], Erica Donner[5], Steven P. Djordjevic[1,2,]

Author affiliations:

[1] Australian Institute for Microbiology & Infection, University of Technology Sydney, NSW, Australia
[2] Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, NSW, Australia
[3] Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries, NSW, Australia
[4] School of Civil, Mining and Environmental Engineering, University of Wollongong, NSW, Australia
[5] Future Industries Institute, University of South Australia, Adelaide, SA, Australia

Table of Contents

Contents

Abbreviations

3`-CS	3`-conserved segment
5`-CS	5`-conserved segment
AAF	Aggregative adherence fimbria
AMR	Antimicrobial resistance
APEC	Avian pathogenic *E. coli*
ARG	Antimicrobial resistance gene
BHC	Bayesian hierarchical clustering
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs
BSI	Bloodstream infection
BWA	Burrows-Wheeler aligner
CARD	Comprehensive antimicrobial resistance database
CC	Clonal complex
CRR	Complex resistance region
DAEC	Diffusely adherent *E. coli*
DHFR	Dihydrofolate reductase
DNA	Desoxyribonucleic acid
EAEC	Enteroaggregative *E. coli*
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ESBL	Extended-spectrum β-lactamase
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal pathogenic *E. coli*
GP	General practitioner
GWA	Genome wide association
HC	Hierarchical cluster
HGT	Horizontal gene transfer
HPI	High-pathogenicity island

Inc	Incompatibility
iTOL	Interactive tree of life
IPEC	Intestinal pathogenic *E. coli*
IS	Insertion sequence
Kb	Kilobases
LB	Lysogeny broth
LEE	Locus of enterocyte effacement
Mb	Megabases
MDR	Multi-drug resistant
MDS	Multidimensional scaling
MGE	Mobile genetic element
MLST	Multi-locus sequence type
MLVA	Multiple locus variable number of tandem repeats analysis
MVC	Melbourne Veterinary Collection
NCBI	National Centre for Biotechnology Information
NMEC	Neonatal meningitis-associated *E. coli*
NSBL	Narrow-spectrum β-lactamase
NSW	New South Wales
OBH	Orange Base Hospital
ORF	Open reading frame
PAI	Pathogenicity-associated island
PCR	Polymerase chain reaction
pMLST	Plasmid multi-locus sequence typing
QRDR	Quinolone resistance determining region
RST	Replicon sequence type
SAE	South Australia Environment
SAH	Sydney Adventist Hospital
SEPEC	Human sepsis-associated *E. coli*
SGI1	Salmonella genomic island 1
SNP	Single nucleotide polymorphism
SPATE	Serine protease autotransporters of *Enterobacteriaceae*

SRA	Sequence read archive
ST	Sequence type
STEC	Shiga toxin-producing *E. coli*
T2SS	Type 2 secretion system
T3SS	Type 3 secretion system
Tn	Transposon
UPEC	Uropathogenic *E. coli*
UTI	Urinary tract infection
VAG	Virulence-associated gene
VFDB	Virulence factor database
WGS	Whole genome sequencing
WHO	World Health Organisation

Abstract

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are the leading cause of urinary tract and bloodstream infections (BSIs) ventilator-associated pneumonia, neonatal meningitis, and wound infections. In Australia, urinary tract infections (UTIs) inflict economic losses of >$AUD 909M annually through ~70,000 hospitalisations and ~2.6 million general practitioner (GP) presentations and are a major driver for antibiotic prescriptions. Here we characterised 426 urine-sourced isolates (circa 2006-2011) from Orange Base Hospital (OBH), a leading rural hospital in New South Wales, Australia. Chapter 4 presents analyses on 336 urine-sourced isolates that were not selected for any phenotypic antibiotic resistances. A total of 91 sequence types (STs) were identified, but pandemic ExPEC lineages ST73, ST127, ST95 and ST131 dominated the collection (146/336). F virulence plasmids carrying *cjrABC-senB* virulence genes cargo were a feature of the collection (present in 15 STs), representing almost 40% (131/336) of the collection whereas UPEC carrying ColV F virulence plasmids (present in 16 STs) were limited in isolation frequency (10.4%; 35 isolates). Class 1 integrons were another feature of the study (n=88; 26.2%). Chapter 5 presents analyses on a collection of trimethoprim resistant isolates (n=67) because trimethoprim has been an important first line antibiotic for treating UTIs for over 40 years. Co-occurrence of trimethoprim-resistance genes with genes encoding extended-spectrum β-lactamases (ESBLs), heavy metals and quaternary ammonium ions was a feature of these ExPEC. ST131 dominated the collection (19/67; 28.4%) of trimethoprim resistant isolates and most carried a class 1 integron. We expanded our study of ST131 to include all isolates with suitable metadata that were of Australian origin. The study, presented in Chapter 6, comprised 284 Australian ST131 *E. coli* isolates from clinical sources, food and companion animals, wildlife, and the environment. Of these, 169 carried a class 1 integron. Forty-one isolates (14 %) contained the complete class one integron-integrase *intI1*. Notably, 128 (45 %) isolates harboured a truncated *intI1* (462-1014 bp). Frequently, these truncations occurred due to the insertion of IS elements, potentially impacting the integrons' capacity to obtain and activate genes. The ST131 H30Rx isolates often carried an antimicrobial gene profile comprising *aac(3)-IIa*, *aac(6)-Ib-cr*, *aph(3')-Ia*, *aadA2*, $bla_{\text{CTX-M-15}}$, $bla_{\text{OXA-1}}$ *and dfrA12*. Notably, dual *parC*-1aAB and *gyrA*-1AB fluoroquinolone resistant mutations were identified in some clade A isolates and this finding requires further monitoring. Although ST131 is an intensively studied ExPEC ST a thorough phylogenomic investigation of ST131 of Australian origin with a One Health perspective was undertaken here. Our study strongly suggests cross-species movement of ST131 strains across diverse reservoirs including humans, gulls, and companion animals.

In conclusion, this study highlights the dominance of pandemic ExPEC lineages ST73, ST127, ST95, and ST131 in urine-sourced isolates from OBH in Australia, and provides a retrospective landscape of AMR in rural NSW which can be used for future comparisons. The presence of F virulence plasmids and class 1 integrons in the collection underscores the importance of molecular surveillance of ExPEC isolates. Moreover, the cross-species movement of ST131 strains across diverse reservoirs underscores the importance of a One Health approach to better understand the epidemiology of ExPEC infections.

Chapter 1: Thesis Overview

1.1. Overview

This thesis, presented by compilation, used whole genome sequencing (WGS) techniques to provide genomic characterisations for 426 *E. coli* isolates (336 – not selected for antibiotics, 67 – trimethoprim resistance selected; 23 – ST131 selected) sourced from urine samples collected from patients attending an Australian rural hospital. The results presented in chapter 5 and chapter 6 were published during candidature and the results presented in chapter 4 are ready for journal submission.

1.2. Aims

This thesis had four major aims:

1) To determine the population structure and multi-locus sequence types (MLST) of a retrospective collection of urine-sourced *E. coli* isolates obtained from an Australian rural hospital using WGS data and phylogenomic analyses.
2) To identify antimicrobial resistance genes (ARGs), virulence-associated genes (VAGs), plasmid replicon sequence types (RSTs) and critical mobile genetic elements (MGEs) (e.g., mobilised complex resistance regions (CRR), insertion sequences (ISs), pathogenicity islands (PAIs), plasmids, phages) using WGS data, multiple databases, read mapping and genotyping tools.
3) Compare isolates selected for trimethoprim resistance to those not selected for phenotypic antibiotic resistance, in terms of populational structure, ARGs, VAGs, and MGEs
4) To determine any statistically significant relationships between host age, sex, or clinical syndrome regarding ARG and VAG profiles

1.3. Summary and knowledge added

This thesis attempts to address the absence of WGS-based studies of UPEC from Australia, particularly from rural and regional settings. It provides comprehensive genomic analyses on two cohorts of urine-sourced *E. coli* from the same hospital – one consisting of isolates selected for phenotypic trimethoprim resistance and the other a large collection of isolates without antibiotic selection criteria. This work also provides the first WGS-based study on Australian ST131 originating from diverse sources (human, animal and environmental).

Some major finding of this thesis include:

- Through phylogenomic analyses, it was determined that antibiotic selection skews population structure. The top three STs in the trimethoprim resistant collection were ST131 > ST69 > ST38. Conversely, in the non-AMR selected collection, the top 3 STs were ST73 > ST95 > ST127.
- The carriage of ARGs conferring resistance to sulfonamide, aminoglycosides, trimethoprim, tetracycline, narrow-spectrum β-lactams (NSBLs), ESBLs and macrolides were substantially higher in isolates selected for trimethoprim resistance compared to non-antibiotic resistance selected isolates, ranging from 1.6 X greater for macrolide resistance genes to 4.5 X greater for ESBL genes. This is likely linked to 74% of the trimethoprim resistant isolates carrying class 1 integrons versus 23% carriage in the non-AMR selected cohort.
- The number and types of ARGs did not differ significantly across patient age, sex, or pathologies; however, isolates that carried *senB*$^{+}$ IncF plasmids that were structurally different to pUTI89 (ExPEC virulence plasmid) had significantly greater ARG totals and class 1 integron integrase *intI1* carriage.
- A strong association with pUTI89 carriage and chromosomal $bla_{\text{CTX-M}}$ genes, particularly in ST131 C1 and A clade isolates.
- No significant differences in number and types of VAGs in terms of patient age and pathologies. Indeed, a genome-wide association (GWA) analysis found no significant differences in VAGs, or any other genes, between UTI and non-UTI-isolates. However, by examining serial UTI patients experiencing worsening sequelae (e.g., cystitis to pyelonephritis) a high proportion of point mutations were found to occur in VAGs.
- There is evidence of potential *E. coli* transmission between different species, as numerous examples of closely related isolates have been identified from both companion and wild animals as well as human isolates, with several OBH isolates (especially those belonging to ST73, ST127, and ST131) showing a genetic similarity of fewer than 100 SNPs to these animal-sourced isolates.
- Similar to other parts of the world, a high portion Australian ST131 population was found to be multi-drug resistant (MDR), though surprisingly 71% of clade A isolates (typically reported as antibiotic sensitive) carried ≥1 ESBL gene. The first report of dual *parC*-1aAB and *gyrA*-1AB mutations (fluoroquinolone resistance) in clade A isolates was also made.

1.4. Short overview of the following chapters:

Chapter 2: This introductory chapter offers a brief overview of the history and genetic structure of E. coli, focusing on the challenges and advancements in classification methods. The emergence of antimicrobial resistance (AMR) within uropathogenic E. coli (UPEC) is a significant concern, and this issue is discussed along with key molecular mechanisms involved. Furthermore, the importance of adopting an OneHealth perspective in addressing these challenges is emphasised.

Chapter 3: This chapter details methods used by the candidate to perform genomic analyses of E. coli isolates. Methods include utilising the Shovill pipeline for sequence data assembly, quality assessment using FastQC and assembly-stats software. Prokka was employed for genome annotation, and various methods, including MLST and single nucleotide polymorphism (SNP)-based alignment, were used to establish phylogenetic relationships. Additionally, gene screening, plasmid analysis, pathogenicity-associated islands (PAI) analysis, and integron analysis were conducted to study MGEs.

Chapter 4: In this results chapter, 336 *E. coli* isolates from patients experiencing UTIs from a regional hospital in Australia were examined. The isolates were not selected based on AMR phenotypes. ST73, ST95, ST127, and ST131 were the dominant STs. F virulence plasmids carrying *senB-cjrABC* genes were found, with distinct characteristics observed for pUTI89-like and non-pUTI89-like plasmids.

Chapter 5: In this results chapter (published), trimethoprim resistant UPEC strains from recurrent UTIs in a rural Australian hospital were sequenced and analysed. The predominant ST was ST131, and we observed co-occurrence of trimethoprim-resistance genes with ESBL genes and genes conferring resistance to heavy metals and quaternary ammonium ions.

Chapter 6: This results chapter (published) provides a detailed phylogenetic analysis of 284 Australian ST131 isolates from various sources. Our findings revealed the typical distribution of ST131 clades A, B, and C, but no specific niche association with some evidence of cross-species movement of strains across different reservoirs. The study identified diverse antibiotic resistance genes (ARGs) and plasmid replicons among the ST131 population in Australia.

Chapter 7: Provides a general discussion. The main aims of this study are to understand the genetic characteristics of UPEC, the impact of antibiotic resistance, and the prevalence of different *E. coli* lineages. Main findings are discussed including the presence of various specific STs in UPEC populations and highlighting the significance of non-resistant lineages and the need

for a comprehensive understanding of the disease burden. The presence of certain ARGs, VAGs, and MGEs in UPEC isolates are also discussed.

Chapter 2: Background

2.1. *Escherichia coli*

E. coli belongs to the Gram-negative *Enterobacteriaceae* family, whose members are widely distributed in nature, and with many living in the gut of humans and other animals (1). Indeed, *E. coli* are commonly isolated from the intestines of warm-blooded organisms but are also found in reptiles (2), fish (3), soil, natural waterbodies, wastewater, and plants (4). Not all genera and species within *Enterobacteriaceae* are pathogenic. Out of 120 *Enterobacteriaceae* species, less than 25 have clinical significance, including *Salmonella* (causing gastroenteritis and BSIs), *Shigella* (causing a plethora of intestinal and systemic infections) and *Yersinia* (causing yersiniosis and bubonic plague) (5). While *E. coli* is a common commensal species, in cases of suppressed immunity or breakage of the gastrointestinal barrier, *E. coli* can cause infection (6). Additionally, *E. coli* can become disease-causing through horizontal gene transfer (HGT) via the acquisition of MGEs such as PAIs, bacteriophages, transposons, and plasmids that contain VAGs (7).

HGT occurs via several mechanisms: transformation, transduction, and conjugation (Figure 2.1). Transformation occurs when desoxyribonucleic acid DNA from one cell is released into the environment, generally via lysis. This DNA can be collected by other bacteria and incorporated into the chromosome or a plasmid. Alternatively, during transduction genomic material is transported from one bacterial cell to another by bacteriophages. Finally, conjugation involves direct contact between two cells with the formation of a mating bridge to transfer plasmids (8).

E. coli have exceptional genome flexibility and rapidly acquire and lose genes via HGT. As such, *E. coli* can adapt to diverse hosts and habitats and is both a common commensal as well the most common Gram-negative bacteria associated with human disease, causing more than 300 million illnesses and about 200,000 deaths annually by diarrhea (9), ~90% of UTIs (10) and is one of the leading cause of sepsis globally (11).

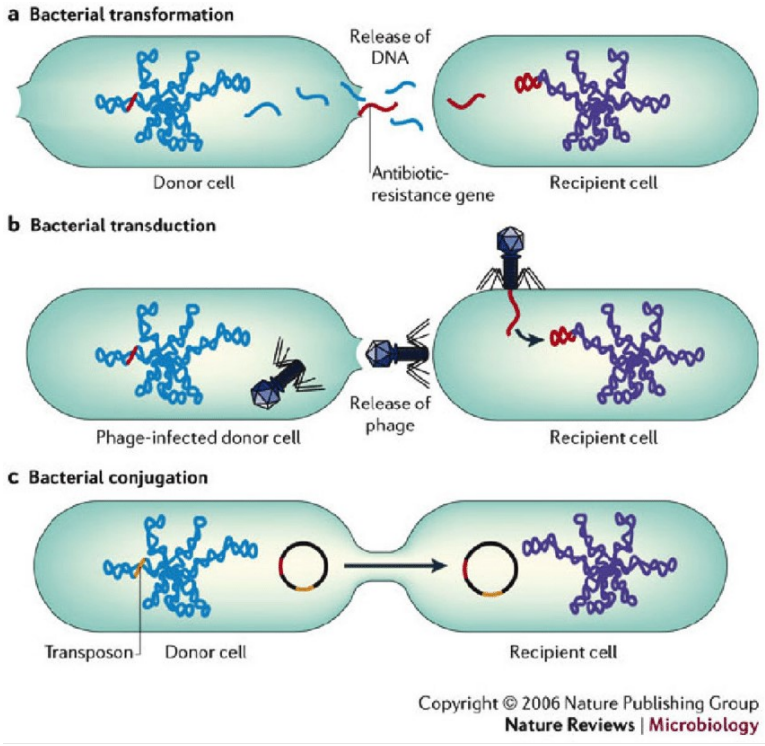


Figure 2.1: **Diagram reviewing the modes of bacterial HGT**. Horizontal transfer can occur when **a)** DNA is released during cell death and obtained by other cells; **b)** by means of bacteriophage; or **c)** conjugation between two bacteria resulting in the exchange of plasmids. Figure from Furuya and Lowy, 2006 (8).

2.1.1. History of *Escherichia coli*

E. coli was first described in 1885 by Theodore Escherich as a Gram-negative bacterium that is commonly found in the lower intestine of warm-blooded organisms, and shortly afterwards it was also isolated from infections outside of gastrointestinal tract (12). Over the next century, the reports of *E. coli* infections increased dramatically (12). One of the most well-known pathogenic *E. coli* is Enterohemorrhagic *E. coli* (EHEC) with serotype O157:H7, which was first identified in 1982 as the cause of an outbreak of severe illness and bloody diarrhea in the United States (13). Early outbreaks of O157:H7 were caused by contaminated meat products, including minced beef and beef burgers (14). Since significant meat hygiene practices were implemented in the late 1980s and early 1990s, there has been a marked reduction in the number of outbreaks associated with these products (14). The largest documented outbreak in the United States, with over 700 illnesses and four deaths, occurred in 1993 in the western United States. The vehicle of transmission was the same as in 1982 — contaminated hamburger, quickly and widely distributed through 'fast food' restaurant chains (15). While the incidences of *E. coli* causing gastrointestinal diseases has been declining, the number of extraintestinal diseases associated with *E. coli*, particularly UTIs and BSIs, has been increasing. For example, a study of 552

cephalosporin-resistant *E. coli* recovered from BSIs in 2014-2015 in Denmark found evidence for 15 national outbreaks (16).

2.1.2. Laboratory detection of *E. coli*

The usual laboratory method for detecting *E. coli* involves culturing the bacteria on selective and differential media that are designed to inhibit the growth of other bacteria while allowing *E. coli* to grow and produce characteristic colonies. These media typically contain compounds such as lactose and pH indicators that allow the identification of *E. coli* based on their ability to ferment lactose and produce acid. Further confirmation of *E. coli* identity can be achieved by performing biochemical tests such as the indole test and the Voges-Proskauer test (17). In addition to traditional culture-based methods, molecular techniques such as PCR (polymerase chain reaction) and MALDI-TOF MS (matrix-assisted laser desorption/ionization – time of flight mass spectroscopy) are increasingly being used for the detection and identification of *E. coli* in clinical settings (17, 18).

2.1.3. Genomic structure of E. coli

E. coli serve as a model organism in molecular biology and genetics, thus the *E. coli* genome has been extensively studied. The first *E. coli* complete genome sequence (strain K-12) was published in 1997 and was comprised of 4.6 million base pairs (bp) and encoded for approximately 4,300 genes (19). The first closed genome of the K-12 strain provided a valuable resource for understanding the molecular biology of bacteria including metabolism and growth, as well as provided tools for identification of potential drug targets for treatment of *E. coli* infections.

E. coli genomes are organised into single, circular chromosomes that range in size from 4.5 to 5.5 Mb (20). *E. coli* genomes, like other bacteria, are further organized into operons, which are clusters of genes that are transcribed together from a single promoter. Operons often encode for functionally related proteins, such as those involved in amino acid biosynthesis or stress response. *E. coli* genomes also contain MGEs, such as plasmids and bacteriophages, the former varying from 3 to 500 kb (21, 22). These elements contribute to the genetic diversity of *E. coli* populations and can play a role in pathogenesis and the spread of antibiotic resistance.

Recombination plays a vital role in the evolutionary dynamics of *E. coli* by enabling the exchange of genetic material between different strains. This process serves as a mechanism for the introduction of novel genetic variations, including virulence factors and ARGs (23, 24). The frequency of recombination is influenced by environment of the strain, with ExPEC isolates exhibiting notably higher rates of recombination compared to their commensal counterparts.

Additionally, there is a positive correlation between the presence of virulence factors and increased recombination frequencies (24). Consequently, recombination events contribute to the emergence of hypervirulent ExPEC lineages, exemplified by the globally disseminated ST131, which is associated with pyelonephritis and AMR (25).

2.2. Classification of *E. coli*

Species classification has been a bedrock in biological work since the Swedish naturalist Carolus Linnaeus introduced the genus species naming system more than 250 years (26). Since then, many classification schemes have been established, particularly for *E. coli,* the most intensively studied organism on Earth (27). These classification systems allow researchers to readily locate and share large amounts of information in a standardised way and provide the information required to identify and track outbreaks or sources of diseases.

2.2.1. Phylogenetic classification of *E. coli*

Phylogenetics refers to the study of evolutionary history and relationships between or within groups of biological entities, usually individuals or genes. Scientists previously relied on colony and cell morphology and techniques such as Gram staining to distinguish *E. coli* from other species. Then with the development of PCR and Sanger sequencing in the late 1970s, it became possible to use proxies like phylogroups and STs (see sections 2.1.3 and 2.1.4 below) to differentiate lineages of *E. coli* further (28). Then, with the development of WGS in the mid-2000s and its increased usage ever since, there has been a shift from phylogenetics to phylogenomics. While phylogenetics usually compares single gene sequences, phylogenomics compares entire genomes, enabling unprecedented precision in studying populational relationships, with strains differentiated by SNPs. Phylogenomics serves as an interface between molecular and evolutionary biology and provides an in-depth understanding on how species evolve via genetic changes, including gene duplications, deletions, and HGT events thereby providing insights into *E. coli* niche adaption and pathogen evolution. For example, *E. coli* ST95 causes disease in humans and poultry; however, these are clearly delineated as distinct lineages on a phylogenetic tree (29, 30). Table 2.1 represents the main pathotypes and serotypes and STs associated with them.

Another way to understand the evolutionary relationships and transmission dynamics of bacterial strains is through SNP typing. By comparing SNPs across different isolates, it is possible to construct phylogenetic trees that reveal the relatedness and ancestry of bacterial populations. In the context of hospital outbreaks, SNP typing becomes particularly crucial for tracking the transmission of pathogens within healthcare settings. A widely accepted threshold

for defining a hospital outbreak is often set at 12-17 SNPs (31-33). This threshold serves as a practical guideline for infection control measures, such as identifying and implementing targeted interventions to prevent further spread. SNP typing is a valuable tool for understanding bacterial strain relatedness, due to its high resolution, which enhance our ability to investigate and control infectious disease spread.

Alternative way of study of relatedness of *E. coli* strains is multiple locus variable number of tandem repeats analysis or MLVA, often used in public health for identification of closely related isolated by comparing their MLVA profiles. This technique is based on variation in the number of repeated DNA sequences at different locations in the genome to distinguish between bacterial strains (34).

MLVA exhibits enhanced cost-effectiveness when compared to WGS. Moreover, MLVA has achieved a well-established status, being extensively employed in numerous research studies and surveillance systems. Conversely, WGS offers significantly heightened resolution and accuracy, concurrently affording valuable insights into evolutionary and phylogenetic aspects, which allowed it to replace MLVA in some outbreak studies (35). Additionally, WGS supplies genomic information that extends beyond the establishment of relationships between isolates, including the identification of VAGs or ARGs. Lastly, WGS data can be easily stored, shared, and subject to reanalysis as new tools or knowledge emerge.

Table 2.1: Main characteristics of the most common *E. coli* pathotypes. From Denamur *et al.* 2021 (36).

Pathotype	Definition basis	Main strain host	Main Virulence genes	Phylogenetic background	Main STs	Main serotypes
ExPEC	Non-intestinal infection, specific genes, animal model	Human, domestic mammals, birds	Genes encoding adhesins, toxins, protectins and iron capture systems	B2	STc131	O16:H5, O25:H4
					STc73	O2:H1, O6:H1
					STc95	O1/O2/O18/O45:K1:H7, O2:K1:H4
					STc12	O4:H1/H5
					STc14	O75:H5
				D	STc69	O17/O73/O77:H18
				C	STc88	O8/O9:H4/H9/H19, O78:H4
				F	STc62	O7:K1:H45
UPEC	Isolated from urine	Human, domestic mammals	*papGII*, *papGIII*	B2	STc131, STc73, STc95, STc12, STc14	Idem ExPEC
				D	STc69	

NMEC	Isolated from cerebrospinal fluid of neonates	Human	Genes encoding the K1 antigen, pS88 genes	B2	STc95	Idem ExPEC
				F	STc59	O1:K1:H7
					STc62	O7:K1:H45
Pneumonia-associated *E. coli*	Isolated from lung	Human	*hly*, *sfa*	B2	STc73	Idem ExPEC
					STc127	O6:H31
					STc141	O2:K1:H6
APEC	Isolated from birds	Poultry	pColV genes	B2	STc95	Idem ExPEC
				C	STc88	O8/O78:H4/H9/H19
				G	STc117	O multiple:H4
IPEC	Diarrhoeal disease	Human, domestic mammals	Various	All phylogroups	Numerous	Numerous
STEC and/or EHEC	*stx* genes	Human, cattle$_c$, sheep$_c$	*stx*, *eae*, *ehxA*	E	STc11	O157:H7
				B1	STc29	O26:H11/H-
						O111:H8/H-
					ST17	O45/O103:H2
EPEC	Attaching and effacing		*eae*, *bfp*	A	ST1788 (EPEC5)	Variable

	lesions on intestinal epithelial cells	Human, domestic mammals			STc10 (EPEC10)	O variable:H40
				B1	STc3 (EPEC2)	O103/O111/O114/
						O126/O128:H2
					STc328 (EPEC7)	O88:H25
						O128/O153/O?:H7
				B2	STc15 (EPEC1)	O55/O127/O142:H6
					STc28 (EPEC4)	O85:H31, O33/O119:H6
					STc5342 (EPEC8)	O55/O76:H51
					STc2346 (EPEC9)	O33/O142:H34
				E	STc335	O55:H7
					STc32	O145:H28
ETEC	Heat-stable and heat-labile enterotoxins	Human, pig, cattle	Genes encoding enterotoxins and colonization factors	A, B1, C, E	Numerous	Numerous

EIEC	Colonocyte invasion	Strictly human	*ipa*, *isc*, *vir*	A	ST6	O124:H30
				B1	ST270	O164:H7
			Inactivation of *nadA*, *nadB* and *cadA*	E	ST280	O143:H26
EAEC	Aggregative adhesion on enterocytes	Human, domestic mammals	Aggregative adherence fimbriae (*aaf/agg*) and transcriptional (*aggR*) genes	A, B1, B2, D	Numerous	Numerous
DAEC	Diffuse adhesion on enterocytes	Human	Genes encoding adhesins (*afa* and *dra*)	All phylogroups	Numerous	Numerous
AIEC	Adhesion and invasion of intestinal epithelial cells	Human	Unknown	All phylogroups with a majority of B2	ST135	O83:H1
					ST73, ST95, ST127, ST131	ExPEC serotypes
Hybrid IPEC	EHEC and EAEC characteristics	Human	*stx*, *aggABCD*, *aggR*	B1	ST678	O104:H4
Hybrid IPEC–ExPEC	HUS and septicaemia	Human, cattle[e]	*stx*, *eae*, pS88 ExPEC genes	A	ST301	O80:H2

2.2.2. Serotyping

Serotyping involves the characterisation of bacteria based on the presence of a few distinct surface antigens (37). Serotypes were discovered at the beginning of the 20th century (37) and have since become a standard for classifying different bacteria, including *E. coli.* The *E. coli* serotyping system is based on three main antigens - O, H and K. There are three known pathways of O-antigen assembly and synthesis, *Wzx/Wzy*, ABC transporter and synthase pathway, and each utilises a particular set of genes for processing of the O antigen (38). Most of the O antigen gene clusters of *E. coli* are located at the *galF-gnd* locus and can be transferred between strains via recombination in the flanking genes, thereby replacing the O antigen (38). The H antigen, or flagellar antigen, is predominantly encoded in the *fliC* locus (39). The flagellum of bacteria projects beyond the cell surface and is rotated to provide movement of the cell (40). Multiple subunits of the protein flagellin compose a single flagellar filament, and the central variable region of the flagellin carries H serotype-specific epitopes (41-44). The K-antigen, which stands for capsule, has a variable presence in *E. coli* strains. There are currently 185 different O-types, 53 H-types (45) and at least 80 K-types (46). Previously, serotyping was only possible *in vitro* using antisera; however, it is now possible to determine serotypes *in silico* using WGS.

2.2.3. Phylogroups (Clermont phylogroups)

A distinct genetic structure in *E. coli* phylogeny was demonstrated in the early 1980s (47, 48). Consequently, a phylogenetic classification scheme, called phylogroups, was created based on the presence and absence of particular genes and gene fragments in isolates located within distinct phylogenetic clusters (49). The original scheme was built to distinguish between four phylogroups - A, B1, B2 and D (49). The original scheme was later updated to include more genes and intergenic regions - from three to eight - thereby enabling the distinction of eight phylogroups – A, B1, B2, C, D, E, F and G, as well as cryptic lineages (50, 51). Phylogroups were initially determined using PCR coupled with gel electrophoresis (49). However, like serotyping, phylogroups can now be determined *in silico* (52). Phylogrouping is a fast and easy way to establish the relatedness between *E. coli* isolates, and strong associations between phylogroups and isolate sources have been demonstrated (53, 54). Figure 2.2 shows phylogenetic history of 72 E. coli selected to represent the phylogenetic diversity of the genus.

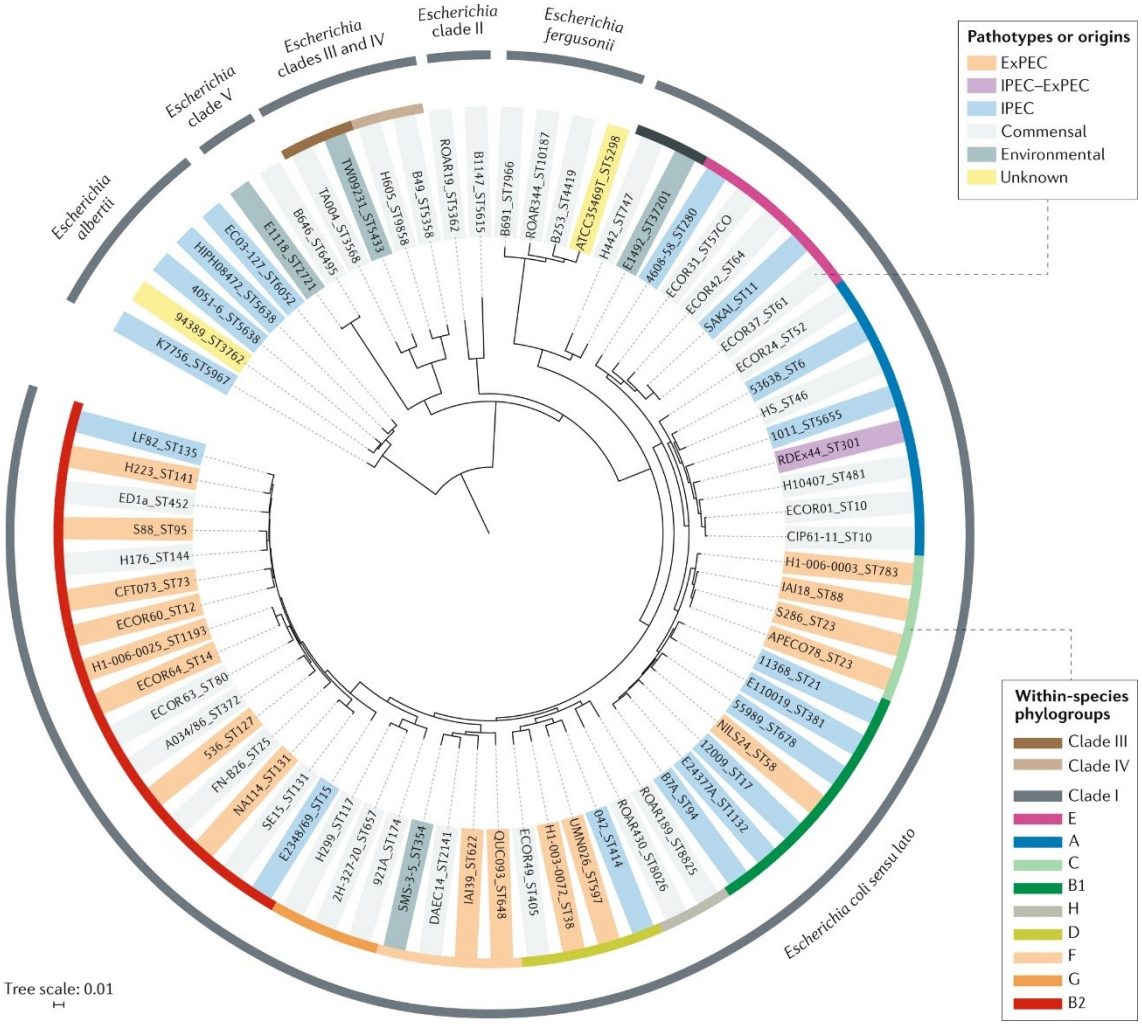


Figure 2.2: Phylogenetic history of *E. coli*. Tree constructed from the core genome of 72 selected isolates to demonstrate the diversity of this genus. From Denamure *et al*. 2021 (36).

2.2.4. Multi-locus sequence typing (MLST)

Compared to phylogrouping, MLST is a more discriminating typing scheme based on the sequences of seven housekeeping gene fragments (50). Isolates with identical gene alleles are assigned to the same ST, while STs that differ in just one or two alleles are grouped into clonal complexes (CCs) (55). MLST can be used in any epidemiological studies in which the bacteria in question exhibits variability in its housekeeping genes and has been applied to tracking the emergence of AMR within lineages and outbreaks caused by novel variants (56). However, there remains a disconnect between the preferred nomenclature used in *E. coli* causing intestinal diseases and *E. coli* causing extraintestinal diseases, with the former typically referred to by serotype (e.g., O157:H7) and the latter by ST (e.g., ST131).

2.2.5. Pathotypes

Pathogenic *E. coli* are broadly categorised as either intestinal pathogenic *E. coli* (IPEC) or ExPEC, and each category comprises multiple pathotypes. *E. coli* pathotypes are based on virulence gene carriage, including those that encode adhesins, invasins, effector molecules, secretion systems, toxins, and siderophore systems. These virulence genes are generally located on MGEs such as phages, PAIs, and plasmids. Plasmids are a diverse group of circular self-replicating extra chromosomal genetic constructs carried and shared by bacteria that often encode AMR, virulence, and fitness genes (57, 58). Plasmids can be classified by incompatibility (Inc) groups based on genes in their replication region and several of the most common Inc groups have a plasmid multi-locus sequence typing (pMLST), enabling greater resolution (59). For greater resolution, plasmids RST schemes also used. RST involves the classification of plasmids based on their replication initiation gene sequences. The method uses a collection of reference replicon sequences to assign plasmids to different replicon types and allows for the identification of plasmids with similar replication origins. This method is particularly useful for studying the epidemiology and transmission of plasmids carrying antibiotic resistance genes and virulence factors (60).

2.2.5.1. Intestinal pathogenic *E. coli* (IPEC)

Multiple *E. coli* pathotypes cause intestinal syndromes in humans, including enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), Shiga toxin-producing *E. coli* (STEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and enteroaggregative *E. coli* (EAEC) (Figure 2.3). EPEC, the first pathotype to be discovered, is an important cause of diarrhoea and premature child death (61). EPEC are characterised by the presence of the Locus of Enterocyte Effacement (LEE) PAI (62-64), containing outer membrane intimin encoded by *eae* gene, type 3 secretory system (T3SS) and multiple T3SS secreted effectors including the translocated receptor for intimin *tir* (65).

EHEC was first described in the 1980s in the US and caused outbreaks of haemorrhagic colitis (66). The main virulence-associated factor of EHEC is the phage-encoded Shiga toxin (*stx*) (67). Epidemiological studies and observation of other Shiga toxin-producing bacteria, including *Shigella dysenteriae,* show that Shiga toxin is responsible for life-threatening conditions like haemorrhagic colitis and haemolytic uraemic syndrome (68, 69). Another important virulence determinant of EHEC is the LEE PAI encoding T3SS and effector proteins, homologous to that produced by EPEC (70). Therefore, isolates producing Shiga toxin but lacking the LEE PAI are referred to as STEC. The most well-known EHEC possess the serotype O157:H7 and are a major foodborne pathogen. *E. coli* O157:H7, and other EHEC lineages, are also known to carry the

pO157 plasmid, which encodes for additional virulence factors, including hemolysin (*ehxA*), catalase-peroxidase (*katP*), type 2 secretion system (*etp*), proteases (*espP* and *stcE*) and adhesin (*toxB*) (71).

ETEC is defined by the ability to produce one or more enterotoxins, including heat-stable (*estIa*) and heat-labile (*eltA*, *eltB*) enterotoxins and their subtypes (72), which are known to be either plasmid (73) or chromosome (within a lambdoid phage encoding region) localised (74). Both toxin types impact osmotic water and electrolyte movement into the intestinal lumen (75). Consequently, ETEC are a leading cause of diarrhoea in children in developing countries, travellers, and in some domestic animals, especially piglets and calves, where it has high economic importance (72, 76-78).

EIEC are genetically and pathogenically similar to *Shigella* spp. While molecular and WGS-based methods can distinguish them (79), EIEC and *Shigella* share many virulence factors and cause similar disease (80). EIEC can cause invasive inflammatory colitis and dysentery (81). EIEC carry the pINV plasmid, which encodes for a T3SS and a set of effectors that allow for the penetration of intestinal epithelial cells, locomotion inside of these cells, and invasion of neighbouring cells (82). EIEC carry other virulence factors like adhesins and toxins, but pINV is the primary virulence determinant restricted to this pathotype (64, 82). By acquiring pINV and other genetic elements, EIEC can survive intracellularly, highlighting the adaptiveness of *E. coli* (83-87).

EAEC is associated with diarrhoea in children from developing countries (88, 89). Originally EAEC was found by a characteristic "stacked-brick" pattern of adherence to cells *in vitro* (90). This behaviour is caused by a type of fimbria known as aggregative adherence fimbria (AAF), often encoded by pAA plasmids (91). EAEC isolates also possess other virulence factors, including the cytotoxin Pet and enteroaggregative stable enterotoxin (EAST-1), which is related to heat-stable enterotoxin of ETEC and a putative enterotoxin ShET1 (64, 88).

DAEC is known to cause diarrhoea in children above age one (92). DAEC are defined by the presence of diffusely adherent pattern on Hep-2 cells, mediated by Afa (*afa*), Dr (*dra*) and F1845 (*daa*) adhesins (81, 93). These adhesins have an affinity for the human decay accelerating factor (CD55) or carcinoembryonic antigen cell adhesion molecules, both located on the intestinal epithelial cells (93). Using Dr fimbriae DAEC strains induce characteristic development of long cellular extensions which wrap around the cell (94).

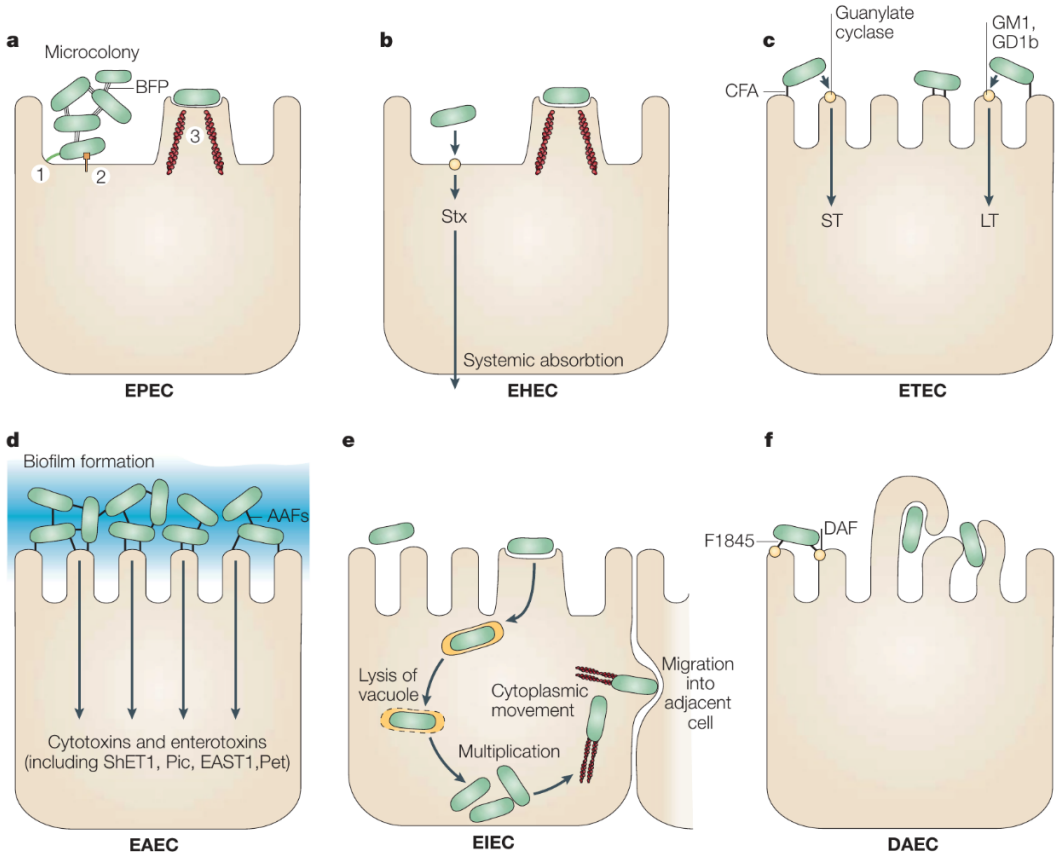


Figure 2.3: Schematic representation of IPEC pathogenesis. **A)** Adherence of EPEC to small bowel enterocyte leading to destruction of normal microvillar architecture. 1. Adhesion, 2. Protein translocation, 3. Formation of the pedestal. **B)** Similarly, EHEC attach and cause effacing lesion, but in the colon. Important characteristic of EHEC is a utilisation of the Shiga toxin which can lead to life-threatening complications. **C)** ETEC attach to small bowel cells and cause watery diarrhoea by the secretion of heat-labile (LT) and/or heat-stable (ST) enterotoxins. **D)** Adherence of EAEC to epithelial cells of small and large bowel in form of a biofilm and secretion of the toxins. **E)** Upon invasion of the epithelial cell, EIEC cells have the ability to lyse the vacuole, generate actin microfilaments to move within the cell, and even migrate between cells. **F)** Characteristic for the DAEC growth of a long cellular extension which wrap around the attached bacteria. AAF, aggregative adherence fimbriae; BFP, bundle-forming pilus; CFA, colonization factor antigen; DAF, decay-accelerating factor; EAST1, EAEC ST1; LT, heat-labile enterotoxin; ShET1, Shigella enterotoxin 1; ST, heat-stable enterotoxin. From Kaper et al., 2004 (95).

2.2.5.2. Extraintestinal pathogenic *E. coli* (ExPEC)

ExPEC are comprised of four main pathotypes: UPEC, which cause UTIs; neonatal meningitis-associated *E. coli* (NMEC); human sepsis-associated *E. coli* (SEPEC), and avian pathogenic *E. coli* (APEC), which cause respiratory infections and septicaemia in poultry (95).

In commensal poultry production, APEC is the etiological agent of colibacillosis, resulting in complex systemic and respiratory infections (96). It was estimated that economic losses associated with APEC in the United States alone reach $40 million annually (97). APEC isolates

utilise multiple virulence-associated factors, including adhesins, invasins, iron acquisition systems and toxins (96). These virulence factors are often found on plasmids and PAIs (98).

Isolates of NMEC pathotype are the most common cause of Gram-negative neonatal meningitis (99). This clinical syndrome has a death rate of 15-40% and can induce severe neurological side defects in survivors, such as neural tube defects (100, 101). Using a neonatal rat model NMEC strains were shown to use S fimbriae to bind to the luminal surfaces of brain microvascular endothelium (102). OmpA and other membrane localised proteins like IbeA and AslA are also required for invasion (103). In addition, an experimental model demonstrated that K1 capsule protein is required for bacterial survival after crossing the blood-brain barrier (104).

SEPEC isolates are associated with bacteraemia and sepsis in patients. SEPEC has a similar virulence profile to NMEC, and includes virulence factors like *ibeA*, *traT*, *iss*, *cvaC* and *sfa*/*foc* (105); however, unlike NMEC, SEPEC tend to carry K2 capsule genes (*kpsMII*) as well as P and F17 fimbriae (106). However, similarities in virulence cargo blur the borders between all ExPEC pathotypes.

The use of animal models has been instrumental in studying ExPEC infections. Animal models, such as mice (107), rats (108) chickens (109) and pigs (110), have provided valuable insights into the pathogenesis (111), virulence factors, host immune response, and potential therapeutics against ExPEC (112). These models allow researchers to simulate and investigate various aspects of ExPEC infections, helping to improve our understanding of the disease and develop effective prevention and treatment strategies.

2.2.5.2.1. Uropathogenic *E. coli* (UPEC)

Uropathogenic *E. coli* are associated with UTIs, an infection of any part of the urinary tract, i.e., urethra, bladder, ureters, or kidneys. UTIs result from a multi-step process of UPEC pathogenesis including: a) Colonisation of the periurethral area with further colonisation of the urethra; b) Ascension to the bladder with growth in a planktonic form in urine; c) Facilitating adherence to the bladder epithelium cells with suppression of cell defence systems; d) Biofilm formation; e) Invasion into host cells and replication by forming bladder intracellular bacterial communities; f) Ascension and colonisation of kidneys with host tissue damage (113). If left untreated, a UTI may lead to urosepsis, which is the spread of the infection into the bloodstream.

UPEC strains possess structural and secretary VAGs contributing to their ability to cause UTIs (113). Before establishing a UTI, UPEC colonise and resist clearance by their host. Colonisation and immune evasion are achieved through VAGs, such as toxins, adhesins and iron acquisition siderophores (Figure 2.4). Common UPEC VAGs include those associated with adherence (Dr

adhesins (*draA-E*), F1C fimbriae (*focA*), P fimbriae (*papG*), S fimbriae (*sfaS, sfaH*), Type 1 fimbriae (*fimH*) and Curlin (*csgA*)), immune evasion (*tcpC*), iron uptake (*iroN,* aerobactin (*iucA-D, iutA*)), proteases (*pic, sat, tsh*) and toxins (*cnf1, hlyA-D*). However, no single VAG profile is a definitive predictor of extra intestinal virulence (114).

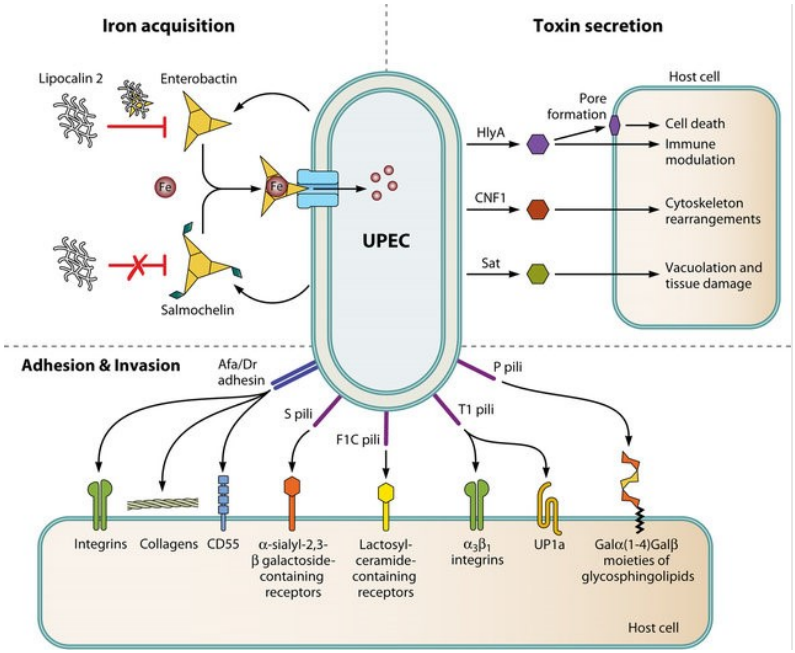


Figure 2.4: Diagram depicting UPEC-associated fitness and VAGs. (Top left) iron acquisition systems, **(top right)** toxins disrupting normal host cell metabolism allowing for easier resource acquisition and immunity evasion, **(bottom)** adhesins and invasins allowing binding to host tissues for example P pili (*pap* operon). From Barber et al., 2016 (115).

UTIs are one of the most common bacterial infections worldwide (116). It is estimated that ~150 million people are diagnosed with UTIs yearly, with UPEC being the most common causative agent and responsible for ~70% of cases (117). MLST-based epidemiological studies have revealed that specific pandemic STs are responsible for more than half of all ExPEC infections globally. These predominant ExPEC lineages include ST131, ST69 (also referred to as 'clonal group A'), ST10, ST405, ST38, ST95, ST73, and ST127 (118, 119). One in two women are diagnosed with UTI in their lifetime, and one in ten women over 18 years old have an episode of UTI every year (120-122). In Australia, UTIs caused 286 hospitalisations per 100,000 people between 2014 and 2015 (123), and the direct economic cost for the Australian healthcare system is estimated to be AU$909 million in the year 2020 with approximately 2.6 million GP presentations (124).

Treatment of UPEC is mostly done empirically and the choice of therapy depends on various factors, including patient age, sex, pregnancy status, history of previous UTIs, and severity of

symptoms. In Australia, trimethoprim, cephalexin, or amoxycillin with clavulanate are used for the majority of acute, uncomplicated infections, in the absence of previous antibiotic exposure or other risk factors such as recent travel to high-risk areas (125). Severe pyelonephritis and sepsis are typically treated with intravenous gentamicin and amoxicillin (126). Nitrofurantoin and fosfomycin are considered second-line and restricted to patients with culture-proven resistant organisms, due to gastrointestinal side effects and low efficacy in patient with poor renal function for nitrofurantoin, while low serum concentration of fosfomycin often lead to treatment failures (127-129).

Furthermore, UTI treatment can be complicated by AMR in the causative agent leading to prolonged hospitalisations and potentially the development of chronic or more severe forms of the disease (130, 131). In particular *E. coli* ST131 has been linked to a worldwide increase in ESBL-producing *E. coli* and consists of three major clades, each associated with specific *fimH* alleles. Clade C, the most prominent subtype identified in human infections (132). *E. coli* ST131 has been linked to a 300% increase in USA hospital admissions due to ESBL-producing *E. coli* in the following decade, driving carbapenem prescription and promoting the spread of potentially untreatable carbapenemase-producing strains (133-135). Therefore, it is vital to track the presence and abundance of ARGs in the UPEC population to provide insights to medical institutions and inform future treatment policies.

2.2.5.3. MGEs associated with ExPEC virulence

The association between UPEC and MGEs is of great interest in understanding the pathogenicity of UPEC strains. Some important UPEC associated MGEs are presented in table 2.2. MGEs, such as plasmids and transposons, play a significant role in the acquisition and dissemination of virulence factors and antibiotic resistance among UPEC isolates.

Table 2.2: Some important MGEs associated with UPEC.

MGE	Approximate size	Important features
pUTI89 (136)	89 kb	Type 1 fimbriae, aerobactin, iron uptake
ColV (98)	140 kb	Salmochelin, aerobactin, colicin V
Tn3 (137)	5 kb	Resistance to ampicillin, encodes transposase for mobility
Tn10 (138)	9 kb	Resistance to tetracycline, exhibits composite transposon structure with ISs
Tn4401 (139)	10 kb	Resistance to carbapenem, contains transposase genes and inverted repeats for transposition
IS1 (140)	0.8 kb	Mediates genomic rearrangements, gene disruption, and promoter capture
IS2 (141)	1.3 kb	Mediates DNA inversions, deletions, and gene fusion
IS3 (142)	1.4 kb	Promotes transposition, gene disruption, and promotes deletions
IS10 (143)	1.5 kb	Involved in gene rearrangements, transposition, and gene disruption
IS26 (144)	0.8 kb	Associated with antibiotic resistance gene capture and mobilization

2.2.5.3.1. ColV plasmids

ColV plasmids are a group of relatively large (usually ~100 Kb) conjugative IncF plasmids, that are thought to be definitive of the APEC pathotype (145). Indeed, ColV plasmids are linked to the ability of APEC to cause colibacillosis in poultry (98, 146), and has been shown to increase the pathogenicity of *E. coli* in a variety of animal models (108, 132, 147). However, the presence of ColV plasmids has also been linked to the ability to cause urosepsis (148) as well as a haemolytic uremic syndrome and neonatal meningitis (108, 133). ColV plasmids are particularly prevalent among certain dominant ExPEC lineages, namely ST95, ST117 and ST58 (30, 146, 149).

ColV plasmids are known to carry a variety of virulence factors like salmochelin (*iroN*), aerobacctin (*iutA*), iron uptake system (*sitA*), colicin V (*cvaA*), increased serum survival (*iss*), type I secretion system (*etsC*), A-haemolysin (*hlyF*) and outer membrane protein (*ompT*), which is important for survival outside of the gastrointestinal tract (Figure 2.5) (98, 150). Additionally,

ColV plasmids often carry class one integrons and their associated ARGs (151). The virulence gene cargo located on ColV plasmids in poultry hosts play important role in colibacillosis and in human hosts they provide advantages to survival outside the gastrointestinal tracts, in particular the urinary tract (152).

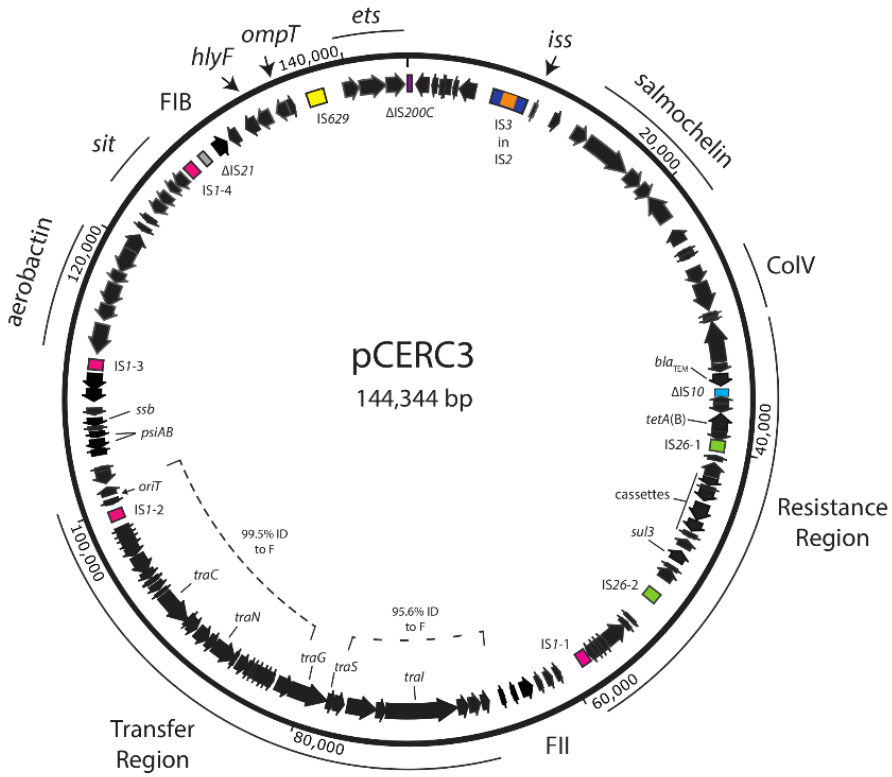


Figure 2.5: ColV plasmid containing resistance region. pCERC3 plasmid with open reading frames (ORFs) shown by arrows and ISs by boxes. VAGs include iron acquisition systems (aerobactin and salmochelin), increased serum survival gene (*iss*) and a hemolysin toxin (*hlyF*). The resistance region contains beta-lactamase bla_{TEM}, tetracycline resistance gene *tetA(B),* and sulphonamide resistance gene *sul3*, as well as a class 1 integron containing a class 1 integron integrase (*intI1*), *dfrA12* (trimethoprim resistance), *aadA2* (aminoglycoside resistance), *cmlA1* (chloramphenicol resistance), *aadA1* (aminoglycoside resistance) and *qacH* (fluoroquinolone resistance). From Moran, Holt and Hall, 2016 (153).

2.2.5.3.2. pUTI89 plasmid

pUTI89 and similar plasmids (like pRS218) are another IncF plasmid, most commonly RST F29:B10 and are ~115 Kbp in size (154, 155). In general, pUTI89 plasmids are divided into two main regions, with one region encoding genes required for conjugative transfer and plasmid replication (i.e., the plasmid "backbone" similar to other incF plasmids) and the other containing virulence-associated and fitness-related genes (Figure 2.6). These genes include enterotoxicity (*senB*), iron acquisition (putative membrane protein and putative periplasmic protein) and copper tolerance genes (*scsC* and *scsD*) (155). Interestingly, three ORFs of pUTI89 are

orthologous to *cjrA*, *cjrB* and *cjrC* genes from EIEC strains and are set in one genomic cluster with enterotoxin SenB (155, 156). pUTI89 plays an important role in the early phase of UTI development and in bladder cell invasion (154, 155). The pUTI89 plasmid is widely present in UPEC strains from cystitis patients (157), and an epidemiological study demonstrated that pUTI89 plasmid is highly prevalent in human-sourced population of *E. coli*, while rarely found in isolates from poultry, pig or cattle (30). These studies suggest that pUTI89 plays an important role in the virulence and/or fitness of human UPEC strains.

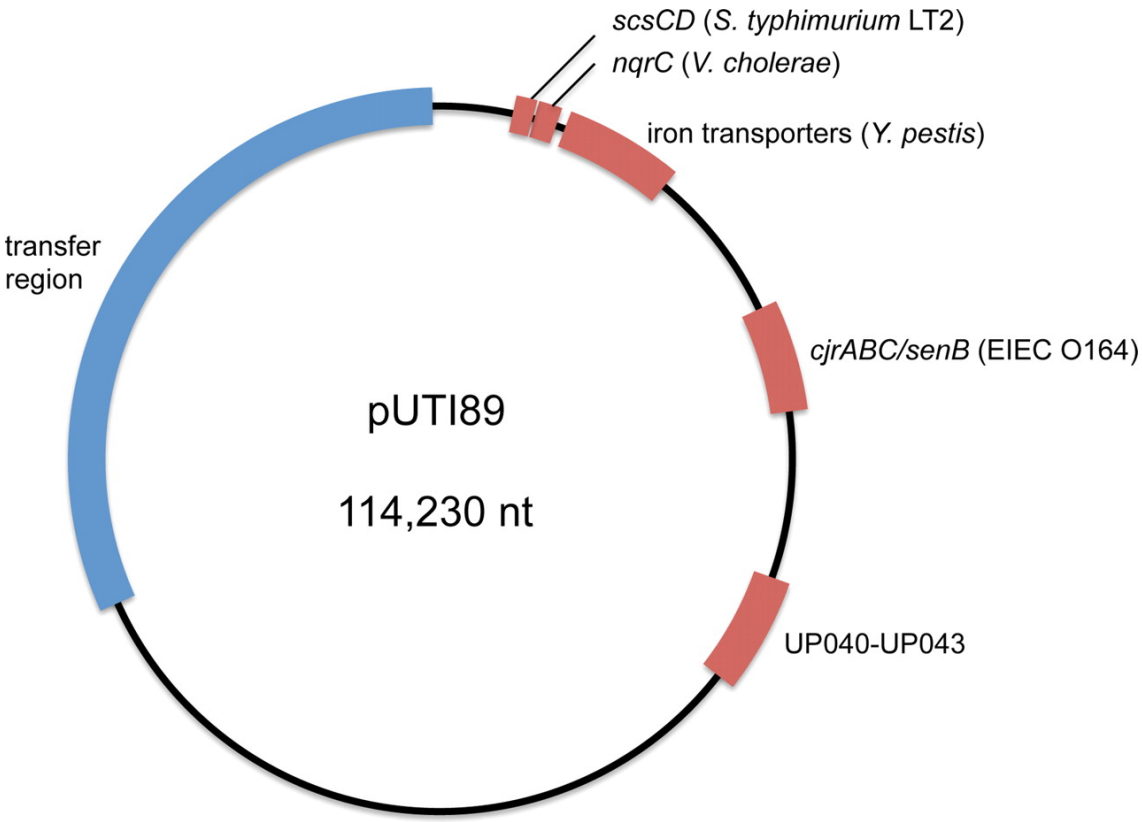


Figure 2.6. Diagram of pUTI89 plasmid. pUTI89 is broadly split into two halves, one (blue) containing a transfer region required for plasmid movement between bacteria, and the other (red) containing known VAGs. Diagram from Cusumano et al. 2010 (155).

2.2.5.3.3. Pathogenicity-associated island (PAI)

Pathogenicity-associated islands are DNA segments, ranging from 10 to 200 Kb, that encode virulence-associated factors including adhesins, invasins, toxins, secretion and iron uptake systems. PAIs are known to reside on chromosomes, plasmids, and bacteriophages. They often have a DNA content that differs from the rest of the host genome, particularly in terms of GC content and codon usage (158). These differences in GC content and codon usage between PAIs and the host genome indicate that the PAI DNA had a different evolutionary history from the rest of the genome and suggest that they have been acquired from a different organism via HGT, or that they have undergone rapid evolution under different selective pressures (159).

The majority of ExPEC virulence factors are encoded by PAIs (160) and studies using the prototype ExPEC strains 536 (161) and CFT073 (162) have shown that PAIs work in unison (163, 164). These studies also demonstrated that virulence increased with each additional PAI, but also highlighted the redundancies of some virulence associated systems (e.g., iron acquisition systems). One of the most well-known PAIs is the *Yersinia* high-pathogenicity island (HPI), which encodes for the siderophore yersiniabactin biosynthesis (165). HPI is widely distributed in the *E. coli* population, present in approximately 40% of fecal isolates and 92% of urosepsis producing isolates (166).

2.2.5.4. Hybrid pathotypes

The introduction of the terms "hetero-pathogens" and "hybrid pathotypes" came with the advent of an outbreak of haemolytic uremic syndrome in Germany in 2011 (167), which was caused by an *E. coli* strain possessing the characteristics of two different pathotypes - EAEC/STEC with O104:H4 serotype (168). This outbreak, which started in Germany and spread to Europe and North America, affected 16 countries (169) and caused ~1000 cases of haemolytic uremic syndrome and 54 deaths (170). The high virulence of this hybrid strain was attributed to the presence Shiga-toxin (*stx*) alongside EAEC associated VAGs, such as aggregative adherence fimbriae I (AAF/I), IrgA homologue adhesin (*iha*) and different autotransporters (like *pic* and *sepA*). These EAEC-associated VAGs allowed for faster colonization and increased absorption of *stx* (171).

Other hybrids have also been documented, including EPEC/ETEC and ExPEC/STEC. The first documented case of EPEC/ETEC hybrid was from six-month-old child with acute watery and bloody diarrhea. The isolate strain carried EPEC-associated *eae* and *elt* genes responsible for attaching and effacing as well as ETEC-associated heat-liable enterotoxin (172). ExPEC/STEC hybrids are concerning as they can simultaneously cause systemic infection and haemolytic uremic syndrome, as seen in an adult patient in France with haemolytic uremic syndrome, diarrhea, and BSI (173). This ExPEC/STEC strain contained Shiga-toxin (*stx2*), intimin (*eae*), enterohaemolysin A (*ehxA*), and a ColV-like plasmid.

While some hybrid pathotypes are known to lead to more severe disease, there is currently insufficient data to ascertain whether hybrid pathogens are always more virulent compared to their parental pathotypes. Therefore, more information about symptomology of patients and genomic data of strains is required to better understand their significance.

2.3. Antimicrobial resistance (AMR)

The World Health Organisation (WHO) has flagged the increasing cases of MDR bacteria worldwide as one of the most serious public health threats (174). As most antibiotics are made of naturally occurring compounds or substances produced by microorganisms, it is unsurprising that, during the course of evolution, bacteria have developed mechanisms to resist these drugs' actions, including the acquisition of AMR genes (175).

The 1950s and 1960s were hailed as the “golden era” of antibiotics as most were discovered during this period (Figure 2.7). However, we have now entered "the discovery void", with no new effective antibiotic classes discovered since 1987 (176, 177). Figure 2.7 also highlights how quickly antibiotic resistance occurs, often occurring within a few years post clinical implementation. AMR is a global threat for a human health, with damaging effect observed all around the world. It is estimated that due to AMR at least 700 000 people die annually (178), with this number could reach 10 million by 2050 (179).

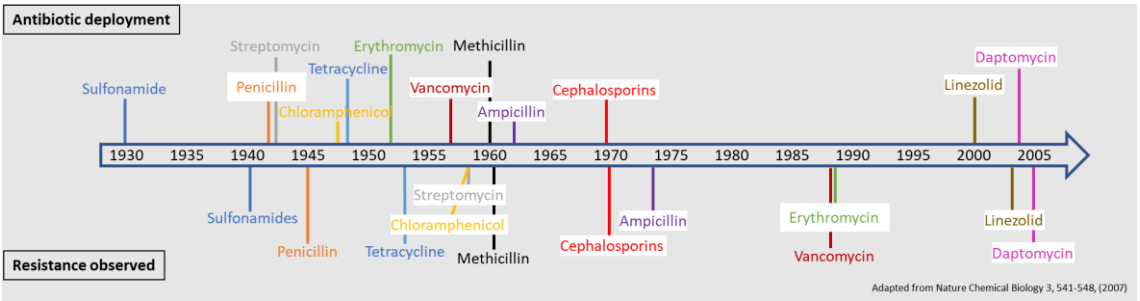


Figure 2.7: **Timeline of new antibiotic discovered vs. resistance observed.** The year, when the antibiotic was deployed, is shown above the timeline, while the year when resistance was observed for the first time is below the timeline. The appearance of resistance rarely exceeds ten years, in cases of erythromycin and vancomycin it took approximately 35 years, however, cephalosporin resistance was detected the next year after deployment. Adapted from Clatworthy et al. 2007 (176).

2.3.1. AMR in *E. coli*

Following introduction of penicillin, evolution of clinically significant resistance to any new antibiotic appeared in as little as a few years (176) (Figure 2.7). However, over the past 20 years, MDR clones within human clinical isolates have been increasing, with the alarming emergence of carbapenemase, and ESBL producing lineages resistant to third and fourth generation cephalosporins and carbapenems (180). Carbapenemase-producing *E. coli* is a major public health threat, as carbapenem was shown to be effective against ESBL-producing *E. coli*. The spread of carbapenem-resistant *E. coli* led to the utilisation of colistin (drug of last resort), especially in food animal production (181). However, in 2015 a mobilised colistin resistance gene

(*mcr-1)* was identified in animals (182) and later in humans (183). The *mcr-1* gene has now spread worldwide, being identified in more than 30 countries (184).

Given that the most common *E. coli* infections are UTIs, and that UTIs are frequently treated via empirical therapy – usually with a prescription of a wide-spectrum antibiotic before any microbiological or antibiotic susceptibility tests are done – this common practice may be a driving factor to increasing MDR abundance in *E. coli* (185).

2.3.1.1. ESBL-producing *E. coli*

ESBL-producing *E. coli* have been placed on the WHO list of critical priority pathogens (186). Genes conferring resistance to ESBLs (*e.g.* $bla_{\text{CTX-M}}$, some bla_{OXA} and bla_{SHV} genes) (187) lead to the inefficacy of many antibiotics from the beta-lactam group, the oldest and most widespread class of antibiotics used clinically (188, 189). Additionally, ESBL resistance often co-occurs with resistance to other classes of antibiotics, like aminoglycosides (e.*g., aadA*, *aph3*) and tetracyclines (*e.g., tetA*, *tetB*) making antibiotic treatment complicated (190). The dominant ESBL positive ST is ST131, a UPEC strain, which was first described in 2008 as a globally spread disease-causing lineage. ST131 isolates commonly produce $bla_{CTX\text{-}M}$ ESBLs (191). Other UPEC ESBL-producing lineages have also emerged, including ST410 and ST648 (192, 193).

2.3.1.2. Trimethoprim resistance

The European Association of Urology guidelines and general European susceptibility data provide the basis for recommending trimethoprim-sulfamethoxazole as an antimicrobial treatment for uncomplicated cystitis and pyelonephritis (194, 195). Similar methods are followed in all parts of the globe (196). However, shortly after the introduction of the drug, resistant *E. coli* strains were reported. The main mechanism behind the resistance was found to be an altered sequence of the dihydrofolate reductase (DHFR) protein (197). Mutation in the *dfr* gene encoding DHFR resulted in a 50 times increase in resistance (198). These mutated *dfr* genes have been captured by multiple MGEs leading to transmissible trimethoprim resistance. Acquired trimethoprim resistance was first reported in the early 1980s (198) and has been increasing ever since (199). Trimethoprim resistance in *E. coli* already reached levels of 31% in the US in 2017 (200), 28% in the UK in 2018 (201) and 24% in Australia in 2019 (202).

2.3.1.3. Fluoroquinolone resistance

Fluoroquinolones are a critically important class of antibiotics for the treatment of serious bacterial infections caused by Gram-positive and Gram-negative bacteria (203, 204). Fluoroquinolones disrupt DNA replication of the bacterial cell by inhibiting two enzymes involved in DNA synthesis and which are absent from human cells (205). Specifically,

fluoroquinolones interfere with the action of DNA gyrase and topoisomerase IV (206). DNA gyrase consists of four subunits encoded by *gyrA* and *gyrB* genes (207, 208) and topoisomerase IV is similarly composed of four subunits encoded by *parC* and *parE* genes (209).

Fluoroquinolones were widely used in human and veterinary medicine because of their effectiveness, but that bears the cost of increased resistance, from 9% in 2015 to 14% in 2019 in Australia alone, while, for example, in the US in 2018-2020 that number reached 22.1% (202). Fluoroquinolone resistance is achieved by two broadly described mechanisms. First are target-site mutations which occur within a short DNA sequence known quinolone resistance determining region (QRDR), especially in *gyrA* (e.g., S83L and D87N) and *parC* (e.g., S80I and E84V). Mutations in the QRDRs of *gyrA* and *parC* can lead to reduced drug binding and decreased susceptibility to quinolones. These mutations often result in amino acid substitutions in critical regions of the proteins, affecting the drug-target interaction and altering the enzymes' ability to carry out their normal functions of DNA replication and repair (210). Second is transmissible quinolone resistance, caused by the acquisition of MGEs containing resistance genes, such as the *qnrA* gene whose product binds to topoisomerase and physically prevents interaction of the antibiotic molecule with the target (211). The rise in fluoroquinolone resistance could be attributed to the high level of global consumption of antibiotic. According to a spatial modelling study published in 2022, the global consumption of fluoroquinolones approximately doubled from 2000 to 2018. The highest consumption rates were observed in eastern Europe and central Asia, followed by Latin America and the Caribbean. The lowest consumption rates were found in sub-Saharan Africa and Oceania including Australia (212, 213). It is probably not a coincidence that the highest level of fluoroquinolone resistance is present in countries with the highest usage (212, 213).

2.3.2. MGE associated with AMR

2.3.2.1. AMR Plasmids

AMR in bacteria is constantly evolving and HGT by plasmids is known to play a primary role. Studying plasmid characteristics and their association with their host is essential for understanding the contribution of plasmids to spread of AMR genes. Out of all known plasmid types, IncF, IncI, IncHI2 and IncX plasmids are known to contribute most to the spread of the AMR in Gram-negative bacteria (114).

Conjugative plasmids of the IncF group are typically low-copy number and have a size range that generally varies from 45 kb to 200 kb. These plasmids have been identified in various bacterial species, including *E. coli*, *Klebsiella pneumoniae*, *Salmonella enterica*, and other

Enterobacteriaceae (214). The ancestral origin of IncF plasmids is thought to be related to the widespread distribution of F plasmids in natural populations of *E. coli*. They are the most common plasmid type to carry ESBL, carbapenemase, aminoglycoside and quinolone resistance genes (215). IncF plasmids are the most prevalent plasmid family in *E. coli*, often providing the bacteria resistance to multiple antimicrobial agents. These plasmids have a system for autonomous replication and encode addiction system, often based on toxin-antitoxin factors, which ensure stable inheritance after cell replication (216, 217). The capture of IncF plasmids carrying ESBL genes, such as $bla_{\text{CTX-M-15}}$, have played an important role in the global dissemination and prevalence of pandemic *E. coli* linages (190).

IncI and IncI2 are also low-copy number, conjugative plasmids, with sizes ranging from 50 to 250 kb (218). IncI plasmids are predominantly described from *E. coli* isolated from poultry and frequently carry ESBL and AmpC genes (215, 219). Additionally, IncI plasmids isolated from *E. coli* CC 10 are associated with ESBL gene $bla_{\text{TEM-52}}$ and livestock (220). IncI2 plasmids on the other hand, are associated with the colistin resistance gene *mcr-1* (221).

IncHI2 plasmids are a subgroup of a IncHI group of plasmids which have size from ~75 to ~400 kb (215). These plasmids have been isolated from human and animal sources, and many are associated with MDR, as they often carry ESBL, sulphonamides, aminoglycosides, tetracyclines and streptomycin resistance determinants (215). Carriage of colistin resistance genes *mcr-1* and *mcr-3* by this plasmid group has also been reported (215, 222).

IncX plasmids are generally smaller compared to other plasmid types, ranging from ~30 to ~50 bp, and there are 6 known subtypes (X1-X6). *E. coli* IncX plasmids are mainly isolated from human and animal sources, and they usually carry ESBL and quinolone resistance genes, but carbapanemases and colistin resistance genes have also been reported (215, 223).

2.3.2.2. Insertion sequences (ISs)

ISs are small MGEs usually carrying one or two transposase genes (224). The sequences of these transposase genes are the primary classification criteria for ISs (225). Some IS elements are known to be tightly associated with AMR in the bacterial genome, as they have tendency to transfer adjacent DNA sequence, which under evolutionary pressure like use of antibiotics can create large resistance gene islands capable of moving between and withing chromosome and plasmid (224). In cases when the insertion is not lethal and under stable positive selective pressure there is a higher probability to maintain AMR associated IS. This insertion site could further facilitate the acquisition of other MGEs and ARGs forming a CRR (226). Probably one of

the most important and widespread is the IS*6* family which includes IS*26* which is known to be associated with different ARGs (227).

2.3.2.3. Transposons

Transposons are DNA elements that encode mechanisms for their intracellular mobility across different locations of the same or different DNA molecules. As they tend to be present in multiple copies on various locations of the genome, they facilitate homologous recombination (228). Specific transposons are known to play role in the spread of multiple ARGs conferring resistance to ampicillin, neomycin, chloramphenicol, and tetracycline, by Tn*3*, Tn*5*, Tn*6* and Tn*9* respectively (229, 230). Another clinically important composite transposon is the Tn*6330* which was found to mobilise the *mcr-1* gene responsible for resistance to the antibiotic of last resort colistin (231). Transposons are generally found on plasmids or within genomic islands in association with ISs. Transposons caring resistance genes for heavy metals, like mercury (e.g., *mer*) and copper (e.g., *cop*) are often associated with ARGs on plasmids due to environmental contamination (232). A great example is Tn*21*, which carries the *mer* operon, and plays a big role in mobilising integrons in natural and clinical environments (233).

2.3.2.4. Integrons

An integron is a small genetic system capable to capture and express genes, in particular genes corresponding to an AMR phenotype (233). Genes captured by integrons are called a cassette array and a single integron can acquire multiple cassettes leading to the MDR phenotype (234). This system protects a host's genome from multiple gene insertions, which can interrupt other biologically important sequences and allows all gene cassettes to be inserted in a particular region (235). Integrons are classified based on the nucleotide sequence of the integrase gene (*intI*) encoding for a site-specific recombinase which controls integration and execution of gene cassettes. Within *E. coli* population, integrons are separated into class 1, 2 and 3, with class 1 integrons being the most clinically important (236). A typical class 1 integron composition includes a 5`-conserved segment (5'-CS) containing *intI1* integrase and variable region at the *attI* site, where gene cassettes are integrated, and a 3`-conserved segment (3'-CS) consisting of a sulfonamide resistance gene (*sul1*) and a truncated quaternary ammonium compound resistance gene (*qacEΔ1*) (237). Class 1 integrons are usually found on plasmids of varying Inc groups and genomic islands in clinical isolates, while environmental isolates usually carry them on the chromosome (238). "Atypical" class 1 integrons on the other hand have part of the 5`-CS and/or 3`-CS missing or altered, like *sul3* instead of *sul1*, or a truncated *intI1* gene often caused by IS elements like IS*26* (239). IS*26*-associated integrons are particularly important in clinical environment, as IS*26*-mobilised genes tend to integrate next to other IS*26* (240).

Gene cassettes are integrated into integrons by an orientation specific insertion, and most cassettes encode ARGs (241, 242). Class 1 integrons are not mobile by nature, they utilise movement machinery of other MGEs, like plasmids, IS and transposons. IS*26*-associated integrons are particularly important in clinical environment, as IS*26*-mobilised genes tend to integrate next to other IS*26* (240). Multiple integron structures have been reported, from different sources and countries, including Australia. However, currently there is only one WGS of integron structure from UTI patients from Australia, and this is limited to just ST73 (243).

2.4. AMR through a One health perspective

Over the past 100 years antibiotics have been used to treat infectious disease, revolutionising modern medicine and food animal production. The main areas where antimicrobials are used are prophylaxis and treatment of human and animal disease, and as a growth promoter in food animal production (244). While antibiotic use in the clinical environment is high, it is estimated that about 70% of antibiotics are used to raise livestock (179). The consequences of the relentless use of these antibiotics are manifold. In the first instance, the overuse has created selection pressures that have led to the creation of bacterial pathogen and commensal populations in humans and animals that are not only resistant to diverse antibiotics but also to metals and disinfectants through co-selection (245). Many antibiotics are poorly absorbed, and a considerable fraction of the administered dose is excreted into the environment (246). Excreted antibiotic residues can then be moved by inappropriate waste management, sewage spillage, direct contact with wildlife, and the use of animal-sourced fertilisers (247, 248), meaning that antibiotic residues provide a constant selection pressure and can affect microbial populations far removed from where the antibiotics were first used. Consequently, clinically important ARGs are not only found in bacteria sourced from humans and food production animals, but also in wildlife, insects, soil (particularly agricultural), natural waterbodies, and air (249-253).

Thus, AMR has been described as the “quintessential One Health issue” (254) and leading regulatory authorities, like the WHO, the International Monetary Fund, and the World Bank, agree that One Health AMR surveillance is vital for identifying AMR reservoirs and mitigating the growing AMR threat to global health (255). As a concept, One Health AMR recognises that the health humans, animals, and the environment are interconnected (Figure 2.8) and requires a multidisciplinary, multisector, and coordinated approach that utilises genomic epidemiology to promote the understanding evolutionary and genetic relationships of AMR in bacteria, hosts and the environment on a global scale (254). One Health acknowledges that socioeconomical issues

influence the emergence of AMR locally, while travel, migration and international trade are big drivers of the global AMR dissemination (255, 256).

E. coli has been proposed as an indicator for One Health AMR due to its omnipresence in humans, animals and the environment. *E. coli* is already an established indicator of faecal contamination, easy to culture, and both healthy and diseased populations can be monitored due to it being a natural resident of the gut microbiome (257). Furthermore, the same antibiotics are often used to treat *E. coli* infections in humans and animals, enabling some comparisons in AMR trends across clinical and agricultural sectors.

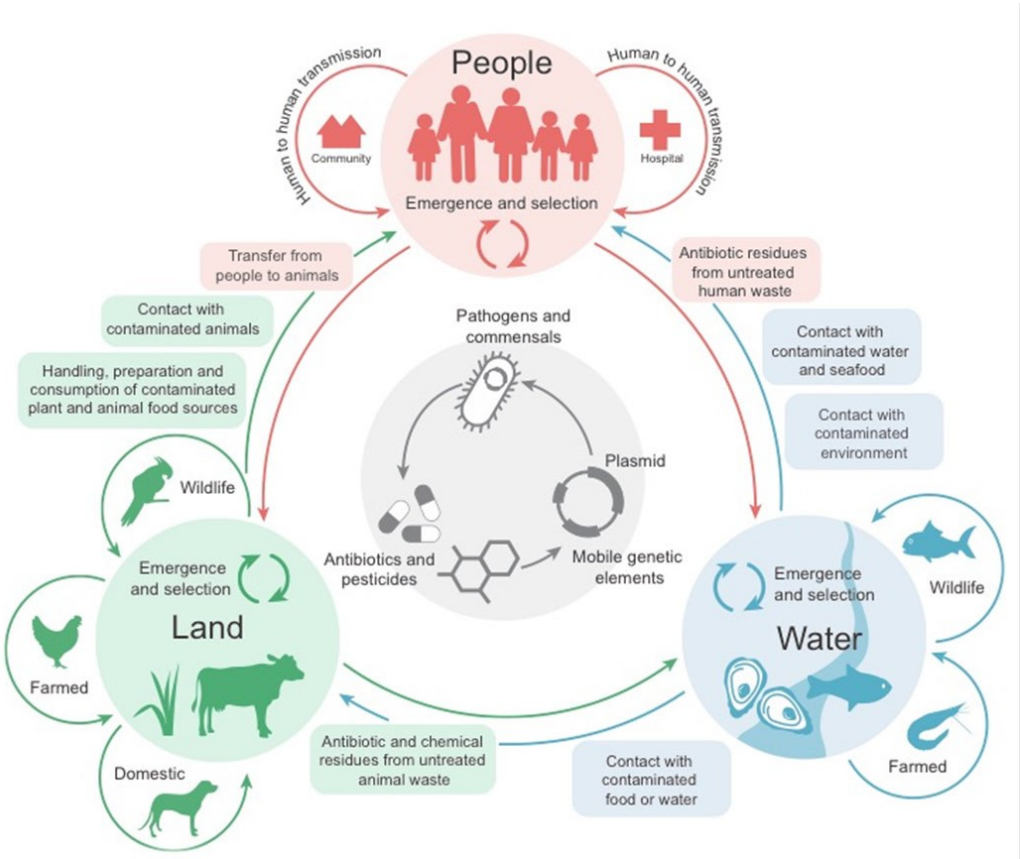


Figure 2.8: Diagram showing the basis of the One Health concept. It highlights the mechanisms involved in the transfer of microorganisms, MGEs, and AMR agents across human, animal and environmental sectors.

2.5. Knowledge gaps

There are currently limited WGS-based studies on UPEC in Australia, particularly in rural and regional settings. Additionally, the existing literature lacks studies on antibiotic selection-free collections and their phylogenetic structure, both within Australia and globally, with the majority of published studies focusing on antibiotic-selected collections, such as ESBL-producing E. coli.

There is also a lack of retrospective genomic characterisation of UPEC collection, including specific STs and their prevalence. Finally, while ST131 is the most reported ExPEC lineage globally, a study on the genetic diversity and evolution of the Australian sourced ST131, has not yet been attempted. The present study aims to address these knowledge gaps by performing WGS study of retrospective antibiotic selection free and trimethoprim resistant collections from rural area of NSW, Australia, as well as comprehensive genomic analysis on all ST131 genomes originating from Australia. This study will determine the phylogenetic structures as well as resistance, virulence and MGE landscape of these three cohorts.

2.6. References

1. Morales-López S, Yepes JA, Prada-Herrera JC, Torres-Jiménez A. Enterobacteria in the 21st century: a review focused on taxonomic changes. J Infect Dev Ctries. 2019;13(4):265-73. doi: 10.3855/jidc.11216.
2. Ramos CP, Santana JA, Morcatti Coura F, Xavier RGC, Leal CAG, Oliveira Junior CA, et al. Identification and Characterization of Escherichia coli, Salmonella Spp., Clostridium perfringens, and C. difficile Isolates from Reptiles in Brazil. Biomed Res Int. 2019;2019:9530732. doi: 10.1155/2019/9530732.
3. Onmaz NE, Yildirim Y, Karadal F, Hizlisoy H, Al S, Gungor C, et al. Escherichia coli O157 in fish: Prevalence, antimicrobial resistance, biofilm formation capacity, and molecular characterization. LWT. 2020;133:109940. doi: 10.1016/j.lwt.2020.109940.
4. NandaKafle G, Christie AA, Vilain S, Brözel VS. Growth and Extended Survival of Escherichia coli O157:H7 in Soil Organic Matter. Front Microbiol. 2018;9:762. doi: 10.3389/fmicb.2018.00762.
5. Rock C, Donnenberg MS. Human Pathogenic Enterobacteriaceae. Reference Module in Biomedical Sciences: Elsevier; 2014.
6. Forde BM, Roberts LW, Phan M-D, Peters KM, Fleming BA, Russell CW, et al. Population dynamics of an Escherichia coli ST131 lineage during recurrent urinary tract infection. Nature Communications. 2019;10(1). doi: 10.1038/s41467-019-11571-5.
7. Bien J, Sokolova O, Bozko P. Role of Uropathogenic Escherichia coli Virulence Factors in Development of Urinary Tract Infection and Kidney Damage. Int J Nephrol. 2012;2012:681473. doi: 10.1155/2012/681473.
8. Furuya EY, Lowy FD. Antimicrobial-resistant bacteria in the community setting. Nat Rev Microbiol. 2006;4(1):36-45. doi: 10.1038/nrmicro1325.
9. Prevention CDC, Brunette GW, Nemhauser JB. CDC Yellow Book 2020: Health Information for International Travel: Oxford University Press; 2019.
10. Poolman JT, Wacker M. Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. The Journal of Infectious Diseases. 2016;213(1):6-13. doi: 10.1093/infdis/jiv429.
11. Organization WH. Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions. 2020.
12. Ørskov I, Ørskov F. Escherichia coli in Extra-Intestinal Infections. The Journal of Hygiene. 1985;95(3):551-75.
13. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. Emerg Infect Dis. 2005;11(4):603-9. doi: 10.3201/eid1104.040739.
14. Pennington TH. E. coli O157 outbreaks in the United Kingdom: past, present, and future. Infect Drug Resist. 2014;7:211-22. doi: 10.2147/idr.S49081.

15. Wachsmuth IK, Sparling PH, Barrett TJ, Potter ME. Enterohemorrhagic Escherichia coli in the United States. FEMS Immunology & Medical Microbiology. 1997;18(4):233-9. doi: 10.1111/j.1574-695X.1997.tb01051.x.
16. Roer L, Hansen F, Thomsen MCF, Knudsen JD, Hansen DS, Wang M, et al. WGS-based surveillance of third-generation cephalosporin-resistant Escherichia coli from bloodstream infections in Denmark. J Antimicrob Chemother. 2017;72(7):1922-9. doi: 10.1093/jac/dkx092.
17. Deisingh AK, Thompson M. Strategies for the detection of Escherichia coli O157:H7 in foods. J Appl Microbiol. 2004;96(3):419-29. doi: 10.1111/j.1365-2672.2003.02170.x.
18. Elabbasy MT, Hussein MA, Algahtani FD, Abd El-Rahman GI, Morshdy AE, Elkafrawy IA, et al. MALDI-TOF MS Based Typing for Rapid Screening of Multiple Antibiotic Resistance E. coli and Virulent Non-O157 Shiga Toxin-Producing E. coli Isolated from the Slaughterhouse Settings and Beef Carcasses. Foods. 2021;10(4). doi: 10.3390/foods10040820.
19. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997;277(5331):1453-62. doi: 10.1126/science.277.5331.1453.
20. Rode CK, Melkerson-Watson LJ, Johnson AT, Bloch CA. Type-specific contributions to chromosome size differences in Escherichia coli. Infect Immun. 1999;67(1):230-6. doi: 10.1128/iai.67.1.230-236.1999.
21. Biet F, Cenatiempo Y, Fremaux C. Identification of a replicon from pTXL1, a small cryptic plasmid from Leuconostoc mesenteroides subsp. mesenteroides Y110, and development of a food-grade vector. Appl Environ Microbiol. 2002;68(12):6451-6. doi: 10.1128/aem.68.12.6451-6456.2002.
22. Kobori H, Sullivan CW, Shizuya H. Bacterial plasmids in antarctic natural microbial assemblages. Appl Environ Microbiol. 1984;48(3):515-8. doi: 10.1128/aem.48.3.515-518.1984.
23. Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, Hoffmann C, et al. Role of intraspecies recombination in the spread of pathogenicity islands within the Escherichia coli species. PLoS Pathog. 2009;5(1):e1000257. doi: 10.1371/journal.ppat.1000257.
24. Rodríguez-Beltrán J, Tourret J, Tenaillon O, López E, Bourdelier E, Costas C, et al. High Recombinant Frequency in Extraintestinal Pathogenic Escherichia coli Strains. Mol Biol Evol. 2015;32(7):1708-16. doi: 10.1093/molbev/msv072.
25. Paul S, Linardopoulou EV, Billig M, Tchesnokova V, Price LB, Johnson JR, et al. Role of homologous recombination in adaptive diversification of extraintestinal Escherichia coli. J Bacteriol. 2013;195(2):231-42. doi: 10.1128/jb.01524-12.
26. Linné Cv. Species plantarum : exhibentes plantas rite cognitas ad genera relatas, cum diferentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas. Berlin: Junk; 1753.
27. Blount ZD. The unexhausted potential of E. coli. Elife. 2015;4. doi: 10.7554/eLife.05826.
28. Young AD, Gillung JP. Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. Systematic Entomology. 2020;45(2):225-47. doi: 10.1111/syen.12406.
29. Donnenberg MS. Escherichia coli : pathotypes and principles of pathogenesis. 2nd edition. ed. Amsterdam: Academic Press; 2013. xxi, 576 pages p.
30. Cummins ML, Reid CJ, Djordjevic SP. F Plasmid Lineages in Escherichia coli ST95: Implications for Host Range, Antibiotic Resistance, and Zoonoses. mSystems. 2022;7(1):e0121221. doi: 10.1128/msystems.01212-21.

31. Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, Colijn C. Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. Mol Biol Evol. 2019;36(3):587-603. doi: 10.1093/molbev/msy242.
32. Ludden C, Coll F, Gouliouris T, Restif O, Blane B, Blackwell GA, et al. Defining nosocomial transmission of Escherichia coli and antimicrobial resistance genes: a genomic surveillance study. Lancet Microbe. 2021;2(9):e472-e80. doi: 10.1016/s2666-5247(21)00117-8.
33. Mills EG, Martin MJ, Luo TL, Ong AC, Maybank R, Corey BW, et al. A one-year genomic investigation of Escherichia coli epidemiology and nosocomial spread at a large US healthcare network. Genome Med. 2022;14(1):147. doi: 10.1186/s13073-022-01150-7.
34. Wang X, Jordan IK, Mayer LW. Chapter 29 - A Phylogenetic Perspective on Molecular Epidemiology. In: Tang Y-W, Sussman M, Liu D, Poxton I, Schwartzman J, editors. Molecular Medical Microbiology (Second Edition). Boston: Academic Press; 2015. p. 517-36.
35. Pijnacker R, van den Beld M, van der Zwaluw K, Verbruggen A, Coipan C, Segura AH, et al. Comparing Multiple Locus Variable-Number Tandem Repeat Analyses with Whole-Genome Sequencing as Typing Method for Salmonella Enteritidis Surveillance in The Netherlands, January 2019 to March 2020. Microbiol Spectr. 2022;10(5):e0137522. doi: 10.1128/spectrum.01375-22.
36. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic Escherichia coli. Nat Rev Microbiol. 2021;19(1):37-54. doi: 10.1038/s41579-020-0416-x.
37. Lancefield RC. A SEROLOGICAL DIFFERENTIATION OF HUMAN AND OTHER GROUPS OF HEMOLYTIC STREPTOCOCCI. J Exp Med. 1933;57(4):571-95. doi: 10.1084/jem.57.4.571.
38. Liu B, Furevi A, Perepelov AV, Guo X, Cao H, Wang Q, et al. Structure and genetics of Escherichia coli O antigens. FEMS Microbiol Rev. 2020;44(6):655-83. doi: 10.1093/femsre/fuz028.
39. Wang L, Rothemund D, Curd H, Reeves PR. Species-wide variation in the Escherichia coli flagellin (H-antigen) gene. J Bacteriol. 2003;185(9):2936-43. doi: 10.1128/jb.185.9.2936-2943.2003.
40. Minamino T, Kinoshita M, Namba K. Directional Switching Mechanism of the Bacterial Flagellar Motor. Comput Struct Biotechnol J. 2019;17:1075-81. doi: 10.1016/j.csbj.2019.07.020.
41. Joys TM. The covalent structure of the phase-1 flagellar filament protein of Salmonella typhimurium and its comparison with other flagellins. J Biol Chem. 1985;260(29):15758-61.
42. Parish CR, Wistar R, Ada GL. Cleavage of bacterial flagellin with cyanogen bromide. Antigenic properties of the protein fragments. Biochem J. 1969;113(3):501-6. doi: 10.1042/bj1130501.
43. Wei LN, Joys TM. Covalent structure of three phase-1 flagellar filament proteins of Salmonella. J Mol Biol. 1985;186(4):791-803. doi: 10.1016/0022-2836(85)90397-3.
44. Winstanley C, Morgan JAW. The bacterial flagellin gene as a biomarker for detection, population genetics and epidemiological analysis. Microbiology (Reading). 1997;143 (Pt 10):3071-84. doi: 10.1099/00221287-143-10-3071.
45. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in Molecular Serotyping and Subtyping of Escherichia coli. Front Microbiol. 2016;7:644. doi: 10.3389/fmicb.2016.00644.
46. 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. J Infect Dis. 2009;199(11):1689-97. doi: 10.1086/598524.

47. Whittam TS, Ochman H, Selander RK. Geographic components of linkage disequilibrium in natural populations of Escherichia coli. Mol Biol Evol. 1983;1(1):67-83. doi: 10.1093/oxfordjournals.molbev.a040302.
48. Whittam TS, Ochman H, Selander RK. Multilocus genetic structure in natural populations of Escherichia coli. Proceedings of the National Academy of Sciences. 1983;80(6):1751-5. doi: doi:10.1073/pnas.80.6.1751.
49. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the Escherichia coli phylogenetic group. Appl Environ Microbiol. 2000;66(10):4555-8. doi: 10.1128/aem.66.10.4555-4558.2000.
50. Clermont O, Dixit OVA, Vangchhia B, Condamine B, Dion S, Bridier-Nahmias A, et al. Characterization and rapid identification of phylogroup G in Escherichia coli, a lineage with high virulence and antibiotic resistance potential. Environ Microbiol. 2019;21(8):3107-17. doi: 10.1111/1462-2920.14713.
51. Clermont O, Gordon DM, Brisse S, Walk ST, Denamur E. Characterization of the cryptic Escherichia lineages: rapid identification and prevalence. Environ Microbiol. 2011;13(9):2468-77. doi: 10.1111/j.1462-2920.2011.02519.x.
52. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. Easy phylotyping of Escherichia coli via the EzClermont web app and command-line tool. Access Microbiol. 2020;2(9):acmi000143. doi: 10.1099/acmi.0.000143.
53. Coura FM, Diniz Sde A, Silva MX, Mussi JM, Barbosa SM, Lage AP, et al. Phylogenetic Group Determination of Escherichia coli Isolated from Animals Samples. ScientificWorldJournal. 2015;2015:258424. doi: 10.1155/2015/258424.
54. Stoppe NC, Silva JS, Carlos C, Sato MIZ, Saraiva AM, Ottoboni LMM, et al. Worldwide Phylogenetic Group Patterns of Escherichia coli from Commensal Human and Wastewater Treatment Plant Isolates. Front Microbiol. 2017;8:2512. doi: 10.3389/fmicb.2017.02512.
55. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. PLoS Pathog. 2012;8(6):e1002776. doi: 10.1371/journal.ppat.1002776.
56. Hastak P, Fourment M, Darling AE, Gottlieb T, Cheong E, Merlino J, et al. Escherichia coli ST8196 is a novel, locally evolved, and extensively drug resistant pathogenic lineage within the ST131 clonal complex. Emerg Microbes Infect. 2020;9(1):1780-92. doi: 10.1080/22221751.2020.1797541.
57. Jain A, Srivastava P. Broad host range plasmids. FEMS Microbiol Lett. 2013;348(2):87-96. doi: 10.1111/1574-6968.12241.
58. Fernández-Alarcón C, Singer RS, Johnson TJ. Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic Escherichia coli from multiple animal sources. PLoS One. 2011;6(8):e23415. doi: 10.1371/journal.pone.0023415.
59. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903. doi: 10.1128/aac.02412-14.
60. Villa L, García-Fernández A, Fortini D, Carattoli A. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. J Antimicrob Chemother. 2010;65(12):2518-29. doi: 10.1093/jac/dkq347.
61. Robins-Browne RM. Traditional enteropathogenic Escherichia coli of infantile diarrhea. Rev Infect Dis. 1987;9(1):28-53. doi: 10.1093/clinids/9.1.28.
62. McDaniel TK, Kaper JB. A cloned pathogenicity island from enteropathogenic Escherichia coli confers the attaching and effacing phenotype on E. coli K-12. Mol Microbiol. 1997;23(2):399-407. doi: 10.1046/j.1365-2958.1997.2311591.x.

63. Robins-Browne RM, Hartland EL. Escherichia coli as a cause of diarrhea. J Gastroenterol Hepatol. 2002;17(4):467-75. doi: 10.1046/j.1440-1746.2002.02769.x.
64. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev. 2013;26(4):822-80. doi: 10.1128/cmr.00022-13.
65. Kenny B, DeVinney R, Stein M, Reinscheid DJ, Frey EA, Finlay BB. Enteropathogenic E. coli (EPEC) transfers its receptor for intimate adherence into mammalian cells. Cell. 1997;91(4):511-20. doi: 10.1016/s0092-8674(00)80437-7.
66. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, et al. Hemorrhagic colitis associated with a rare Escherichia coli serotype. N Engl J Med. 1983;308(12):681-5. doi: 10.1056/nejm198303243081203.
67. Melton-Celsa A, Mohawk K, Teel L, O'Brien A. Pathogenesis of Shiga-toxin producing escherichia coli. Curr Top Microbiol Immunol. 2012;357:67-103. doi: 10.1007/82_2011_176.
68. Walker CL, Applegate JA, Black RE. Haemolytic-uraemic syndrome as a sequela of diarrhoeal disease. J Health Popul Nutr. 2012;30(3):257-61. doi: 10.3329/jhpn.v30i3.12288.
69. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, et al. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N Engl J Med. 2011;365(8):718-24. doi: 10.1056/NEJMoa1107643.
70. Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. Parallel evolution of virulence in pathogenic Escherichia coli. Nature. 2000;406(6791):64-7. doi: 10.1038/35017546.
71. Lim JY, Yoon J, Hovde CJ. A brief overview of Escherichia coli O157:H7 and its plasmid O157. J Microbiol Biotechnol. 2010;20(1):5-14.
72. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev. 2005;18(3):465-83. doi: 10.1128/cmr.18.3.465-483.2005.
73. Echeverria P, Seriwatana J, Taylor DN, Changchawalit S, Smyth CJ, Twohig J, et al. Plasmids coding for colonization factor antigens I and II, heat-labile enterotoxin, and heat-stable enterotoxin A2 in Escherichia coli. Infect Immun. 1986;51(2):626-30. doi: 10.1128/iai.51.2.626-630.1986.
74. Jobling MG. The chromosomal nature of LT-II enterotoxins solved: a lambdoid prophage encodes both LT-II and one of two novel pertussis-toxin-like toxin family members in type II enterotoxigenic Escherichia coli. Pathog Dis. 2016;74(3). doi: 10.1093/femspd/ftw001.
75. Read LT, Hahn RW, Thompson CC, Bauer DL, Norton EB, Clements JD. Simultaneous exposure to Escherichia coli heat-labile and heat-stable enterotoxins increases fluid secretion and alters cyclic nucleotide and cytokine production by intestinal epithelial cells. Infect Immun. 2014;82(12):5308-16. doi: 10.1128/iai.02496-14.
76. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. PLoS One. 2013;8(9):e72788. doi: 10.1371/journal.pone.0072788.
77. Nagy B, Fekete PZ. Enterotoxigenic Escherichia coli (ETEC) in farm animals. Vet Res. 1999;30(2-3):259-84.
78. Fairbrother JM, Nadeau E, Gyles CL. Escherichia coli in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. Anim Health Res Rev. 2005;6(1):17-39. doi: 10.1079/ahr2005105.
79. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and

strategies. New Microbes New Infect. 2018;21:58-62. doi: 10.1016/j.nmni.2017.09.003.
80. BRENNER DJ, Fanning G, Miklos G, Steigerwalt A. Polynucleotide sequence relatedness among Shigella species. International Journal of Systematic and Evolutionary Microbiology. 1973;23(1):1-7.
81. Nataro JP, Kaper JB. Diarrheagenic Escherichia coli. Clin Microbiol Rev. 1998;11(1):142-201. doi: 10.1128/cmr.11.1.142.
82. Marteyn B, Gazi A, Sansonetti P. Shigella: a model of virulence regulation in vivo. Gut Microbes. 2012;3(2):104-20. doi: 10.4161/gmic.19325.
83. Day WA, Jr., Fernández RE, Maurelli AT. Pathoadaptive mutations that enhance virulence: genetic organization of the cadA regions of Shigella spp. Infect Immun. 2001;69(12):7471-80. doi: 10.1128/iai.69.12.7471-7480.2001.
84. Feng Y, Chen Z, Liu SL. Gene decay in Shigella as an incipient stage of host-adaptation. PLoS One. 2011;6(11):e27754. doi: 10.1371/journal.pone.0027754.
85. Prosseda G, Di Martino ML, Campilongo R, Fioravanti R, Micheli G, Casalino M, et al. Shedding of genes that interfere with the pathogenic lifestyle: the Shigella model. Res Microbiol. 2012;163(6-7):399-406. doi: 10.1016/j.resmic.2012.07.004.
86. Prunier AL, Schuch R, Fernández RE, Mumy KL, Kohler H, McCormick BA, et al. nadA and nadB of Shigella flexneri 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of Shigella pathogenicity. Microbiology (Reading). 2007;153(Pt 7):2363-72. doi: 10.1099/mic.0.2007/006916-0.
87. Amhaz JM, Andrade A, Bando SY, Tanaka TL, Moreira-Filho CA, Martinez MB. Molecular typing and phylogenetic analysis of enteroinvasive Escherichia coli using the fliC gene sequence. FEMS Microbiol Lett. 2004;235(2):259-64. doi: 10.1016/j.femsle.2004.04.044.
88. Okeke IN, Nataro JP. Enteroaggregative Escherichia coli. Lancet Infect Dis. 2001;1(5):304-13. doi: 10.1016/s1473-3099(01)00144-x.
89. Harrington SM, Dudley EG, Nataro JP. Pathogenesis of enteroaggregative Escherichia coli infection. FEMS Microbiol Lett. 2006;254(1):12-8. doi: 10.1111/j.1574-6968.2005.00005.x.
90. Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, Levine MM. Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. Pediatr Infect Dis J. 1987;6(9):829-31. doi: 10.1097/00006454-198709000-00008.
91. Boisen N, Hansen AM, Melton-Celsa AR, Zangari T, Mortensen NP, Kaper JB, et al. The presence of the pAA plasmid in the German O104:H4 Shiga toxin type 2a (Stx2a)-producing enteroaggregative Escherichia coli strain promotes the translocation of Stx2a across an epithelial cell monolayer. J Infect Dis. 2014;210(12):1909-19. doi: 10.1093/infdis/jiu399.
92. Scaletsky IC, Fabbricotti SH, Carvalho RL, Nunes CR, Maranhão HS, Morais MB, et al. Diffusely adherent Escherichia coli as a cause of acute diarrhea in young children in Northeast Brazil: a case-control study. J Clin Microbiol. 2002;40(2):645-8. doi: 10.1128/jcm.40.2.645-648.2002.
93. Servin AL. Pathogenesis of human diffusely adhering Escherichia coli expressing Afa/Dr adhesins (Afa/Dr DAEC): current insights and future challenges. Clin Microbiol Rev. 2014;27(4):823-69. doi: 10.1128/cmr.00036-14.
94. Bernet-Camard MF, Coconnier MH, Hudault S, Servin AL. Pathogenicity of the diffusely adhering strain Escherichia coli C1845: F1845 adhesin-decay accelerating factor interaction, brush border microvillus injury, and actin disassembly in cultured human intestinal epithelial cells. Infect Immun. 1996;64(6):1918-28. doi: 10.1128/iai.64.6.1918-1928.1996.

95. Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. Nat Rev Microbiol. 2004;2(2):123-40. doi: 10.1038/nrmicro818.
96. Dziva F, Stevens MP. Colibacillosis in poultry: unravelling the molecular basis of virulence of avian pathogenic Escherichia coli in their natural hosts. Avian Pathol. 2008;37(4):355-66. doi: 10.1080/03079450802216652.
97. de Brito BG, Gaziri LC, Vidotto MC. Virulence factors and clonal relationships among Escherichia coli strains isolated from broiler chickens with cellulitis. Infect Immun. 2003;71(7):4175-7. doi: 10.1128/iai.71.7.4175-4177.2003.
98. Johnson TJ, Siek KE, Johnson SJ, Nolan LK. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian Escherichia coli strains. J Bacteriol. 2006;188(2):745-58. doi: 10.1128/jb.188.2.745-758.2006.
99. Wijetunge DS, Gongati S, DebRoy C, Kim KS, Couraud PO, Romero IA, et al. Characterizing the pathotype of neonatal meningitis causing Escherichia coli (NMEC). BMC Microbiol. 2015;15:211. doi: 10.1186/s12866-015-0547-9.
100. Unhanand M, Mustafa MM, McCracken GH, Jr., Nelson JD. Gram-negative enteric bacillary meningitis: a twenty-one-year experience. J Pediatr. 1993;122(1):15-21. doi: 10.1016/s0022-3476(05)83480-8.
101. Dawson KG, Emerson JC, Burns JL. Fifteen years of experience with bacterial meningitis. Pediatr Infect Dis J. 1999;18(9):816-22. doi: 10.1097/00006454-199909000-00014.
102. Parkkinen J, Korhonen TK, Pere A, Hacker J, Soinila S. Binding sites in the rat brain for Escherichia coli S fimbriae associated with neonatal meningitis. J Clin Invest. 1988;81(3):860-5. doi: 10.1172/jci113395.
103. Kim KS. Escherichia coli translocation at the blood-brain barrier. Infect Immun. 2001;69(9):5217-22. doi: 10.1128/iai.69.9.5217-5222.2001.
104. Hoffman JA, Wass C, Stins MF, Kim KS. The capsule supports survival but not traversal of Escherichia coli K1 across the blood-brain barrier. Infect Immun. 1999;67(7):3566-70. doi: 10.1128/iai.67.7.3566-3570.1999.
105. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev. 2004;17(1):14-56. doi: 10.1128/cmr.17.1.14-56.2004.
106. Jakobsen L, Garneau P, Kurbasic A, Bruant G, Stegger M, Harel J, et al. Microarray-based detection of extended virulence and antimicrobial resistance gene profiles in phylogroup B2 Escherichia coli of human, meat and animal origin. J Med Microbiol. 2011;60(Pt 10):1502-11. doi: 10.1099/jmm.0.033993-0.
107. Jakobsen L, Hammerum AM, Frimodt-Møller N. Detection of clonal group A Escherichia coli isolates from broiler chickens, broiler chicken meat, community-dwelling humans, and urinary tract infection (UTI) patients and their virulence in a mouse UTI model. Appl Environ Microbiol. 2010;76(24):8281-4. doi: 10.1128/aem.01874-10.
108. 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. doi: 10.1128/iai.00347-10.
109. Moulin-Schouleur M, Répérant M, Laurent S, Brée A, Mignon-Grasteau S, Germon P, et al. Extraintestinal pathogenic Escherichia coli strains of avian and human origin: link between phylogenetic relationships and common virulence patterns. J Clin Microbiol. 2007;45(10):3366-76. doi: 10.1128/jcm.00037-07.
110. Predojević L, Keše D, Žgur Bertok D, Železnik Ramuta T, Veranič P, Erdani Kreft M, et al. A Biomimetic Porcine Urothelial Model for Assessing Escherichia coli Pathogenicity. Microorganisms. 2022;10(4). doi: 10.3390/microorganisms10040783.
111. Kromann S, Jensen HE. In vivo models of Escherichia coli infection in poultry. Acta Vet Scand. 2022;64(1):33. doi: 10.1186/s13028-022-00652-z.

112. Cox E, Aloulou M, Fleckenstein JM, Schäffer C, Sjöling Å, Schüller S, et al. The Intriguing Interaction of Escherichia coli with the Host Environment and Innovative Strategies To Interfere with Colonization: a Summary of the 2019 E. coli and the Mucosal Immune System Meeting. Appl Environ Microbiol. 2020;86(24). doi: 10.1128/aem.02085-20.
113. Terlizzi ME, Gribaudo G, Maffei ME. UroPathogenic Escherichia coli (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. Front Microbiol. 2017;8:1566. doi: 10.3389/fmicb.2017.01566.
114. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019;47(D1):D687-d92. doi: 10.1093/nar/gky1080.
115. Barber AE, Norton JP, Wiles TJ, Mulvey MA. Strengths and Limitations of Model Systems for the Study of Urinary Tract Infections and Related Pathologies. Microbiol Mol Biol Rev. 2016;80(2):351-67. doi: 10.1128/mmbr.00067-15.
116. Stamm WE, Norrby SR. Urinary tract infections: disease panorama and challenges. J Infect Dis. 2001;183 Suppl 1:S1-4. doi: 10.1086/318850.
117. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol. 2015;13(5):269-84. doi: 10.1038/nrmicro3432.
118. 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). doi: 10.1128/cmr.00135-18.
119. Yamaji R, Rubin J, Thys E, Friedman CR, Riley LW. Persistent Pandemic Lineages of Uropathogenic Escherichia coli in a College Community from 1999 to 2017. J Clin Microbiol. 2018;56(4). doi: 10.1128/jcm.01834-17.
120. Micali S, Isgro G, Bianchi G, Miceli N, Calapai G, Navarra M. Cranberry and recurrent cystitis: more than marketing? Crit Rev Food Sci Nutr. 2014;54(8):1063-75. doi: 10.1080/10408398.2011.625574.
121. Foxman B, Brown P. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect Dis Clin North Am. 2003;17(2):227-41. doi: 10.1016/s0891-5520(03)00005-9.
122. Foxman B. Urinary Tract Infection Syndromes: Occurrence, Recurrence, Bacteriology, Risk Factors, and Disease Burden. Infectious Disease Clinics of North America. 2014;28(1):1-13. doi: https://doi.org/10.1016/j.idc.2013.09.003.
123. AURA 2016: first Australian report on antimicrobial use and resistance in human health – summary report: Australian Commission on Safety and Quality in Health Care (ACSQHC) Sydney: ACSQHC; 2016 [
124. A One Health antimicrobial resistance economic perspective: OUTBREAK consortium, University of Technology Sydney, NSW 2007, Australia; 2020 [
125. Jarvis TR, Chan L, Gottlieb T. Assessment and management of lower urinary tract infection in adults. Aust Prescr. 2014;37(1):7-9. doi: 10.18773/austprescr.2014.002.
126. Alam M, Bastakoti B. Therapeutic Guidelines: Antibiotic. Version 15. Aust Prescr. 2015;38(4). doi: 10.18773/austprescr.2015.049.
127. Gardiner BJ, Stewardson AJ, Abbott IJ, Peleg AY. Nitrofurantoin and fosfomycin for resistant urinary tract infections: old drugs for emerging problems. Aust Prescr.42(1):14-9. doi: 10.18773/austprescr.2019.002.
128. Somorin YM, Weir NM, Pattison SH, Crockard MA, Hughes CM, Tunney MM, et al. Antimicrobial resistance in urinary pathogens and culture-independent detection of trimethoprim resistance in urine from patients with urinary tract infection. BMC Microbiol. 2022;22(1):144. doi: 10.1186/s12866-022-02551-9.

129. Gardiner BJ, Stewardson AJ, Abbott IJ, Peleg AY. Nitrofurantoin and fosfomycin for resistant urinary tract infections: old drugs for emerging problems. Aust Prescr. 2019;42(1):14-9. doi: 10.18773/austprescr.2019.002.
130. Mulvey MA, Schilling JD, Hultgren SJ. Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. Infect Immun. 2001;69(7):4572-9. doi: 10.1128/iai.69.7.4572-4579.2001.
131. Wagenlehner FME, Pilatz A, Weidner W, Naber KG. Urosepsis: Overview of the Diagnostic and Treatment Challenges. Microbiol Spectr. 2015;3(5). doi: 10.1128/microbiolspec.UTI-0003-2012.
132. 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. doi: 10.1128/iai.00363-06.
133. Cointe A, Birgy A, Mariani-Kurkdjian P, Liguori S, Courroux C, Blanco J, et al. Emerging Multidrug-Resistant Hybrid Pathotype Shiga Toxin-Producing Escherichia coli O80 and Related Strains of Clonal Complex 165, Europe. Emerg Infect Dis. 2018;24(12):2262-9. doi: 10.3201/eid2412.180272.
134. Saidenberg ABS, Stegger M, Price LB, Johannesen TB, Aziz M, Cunha MPV, et al. mcr-Positive Escherichia coli ST131-H22 from Poultry in Brazil. Emerg Infect Dis. 2020;26(8):1951-4. doi: 10.3201/eid2608.191724.
135. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, et al. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res. 2017;27(8):1437-49. doi: 10.1101/gr.216606.116.
136. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence Plasmid Harbored by Uropathogenic Escherichia coli Functions in Acute Stages of Pathogenesis. Infection and Immunity. 2010;78(4):1457-67. doi: 10.1128/iai.01260-09.
137. Gerlach BA, Wiedemann B. Tn3 as the molecular basis of ampicillin resistance in E. coli--an epidemiological survey. Zentralbl Bakteriol Mikrobiol Hyg A. 1985;260(1):139-50. doi: 10.1016/s0176-6724(85)80110-3.
138. Billips BK, Schaeffer AJ, Klumpp DJ. Molecular basis of uropathogenic Escherichia coli evasion of the innate immune response in the bladder. Infect Immun. 2008;76(9):3891-900. doi: 10.1128/iai.00069-08.
139. Cuzon G, Naas T, Nordmann P. Functional characterization of Tn4401, a Tn3-based transposon involved in blaKPC gene mobilization. Antimicrob Agents Chemother. 2011;55(11):5370-3. doi: 10.1128/aac.05202-11.
140. Biel SW, Berg DE. Mechanism of IS1 transposition in E. coli: choice between simple insertion and cointegration. Genetics. 1984;108(2):319-30. doi: 10.1093/genetics/108.2.319.
141. Goyard S, Pidoux J, Ullmann A. An Escherichia coli insertion element (IS2) provides a functional promoter in Bordetella pertussis. Res Microbiol. 1991;142(6):633-41. doi: 10.1016/0923-2508(91)90076-m.
142. Charlier D, Piette J, Glansdorff N. IS3 can function as a mobile promoter in E. coli. Nucleic Acids Res. 1982;10(19):5935-48. doi: 10.1093/nar/10.19.5935.
143. Harayama S, Oguchi T, Iino T. The E. coli K-12 chromosome flanked by two IS10 sequences transposes. Mol Gen Genet. 1984;197(1):62-6. doi: 10.1007/bf00327923.
144. Roy Chowdhury P, McKinnon J, Liu M, Djordjevic SP. Multidrug Resistant Uropathogenic Escherichia coli ST405 With a Novel, Composite IS26 Transposon in a Unique Chromosomal Location. Front Microbiol. 2018;9:3212. doi: 10.3389/fmicb.2018.03212.

145. 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. doi: 10.1128/jcm.00816-08.
146. 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. Microb Genom. 2019;5(2). doi: 10.1099/mgen.0.000250.
147. 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 (Reading). 2009;155(Pt 2):450-60. doi: 10.1099/mic.0.023143-0.
148. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. Int J Antimicrob Agents. 2018;52(3):430-5. doi: 10.1016/j.ijantimicag.2018.06.017.
149. Reid CJ, Cummins ML, Börjesson S, Brouwer MSM, Hasman H, Hammerum AM, et al. A role for ColV plasmids in the evolution of pathogenic Escherichia coli ST58. Nat Commun. 2022;13(1):683. doi: 10.1038/s41467-022-28342-4.
150. 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. doi: 10.1089/fpd.2011.0961.
151. 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. doi: 10.1089/mdr.2017.0177.
152. 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 (Reading). 2005;151(Pt 6):2097-110. doi: 10.1099/mic.0.27499-0.
153. Moran RA, Holt KE, Hall RM. pCERC3 from a commensal ST95 Escherichia coli: A ColV virulence-multiresistance plasmid carrying a sul3-associated class 1 integron. Plasmid. 2016;84-85:11-9. doi: 10.1016/j.plasmid.2016.02.002.
154. Wijetunge DS, Karunathilake KH, Chaudhari A, Katani R, Dudley EG, Kapur V, et al. Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated Escherichia coli strain RS218. BMC Microbiol. 2014;14:203. doi: 10.1186/s12866-014-0203-9.
155. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect Immun. 2010;78(4):1457-67. doi: 10.1128/iai.01260-09.
156. Huang WC, Liao YJ, Hashimoto M, Chen KF, Chu C, Hsu PC, et al. cjrABC-senB hinders survival of extraintestinal pathogenic E. coli in the bloodstream through triggering complement-mediated killing. J Biomed Sci. 2020;27(1):86. doi: 10.1186/s12929-020-00677-4.
157. DebRoy C, Sidhu MS, Sarker U, Jayarao BM, Stell AL, Bell NP, et al. Complete sequence of pEC14_114, a highly conserved IncFIB/FIIA plasmid associated with uropathogenic Escherichia coli cystitis strains. Plasmid. 2010;63(1):53-60. doi: 10.1016/j.plasmid.2009.10.003.
158. Novick RP. Pathogenicity and Other Genomic Islands. In: Maloy S, Hughes K, editors. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press; 2013. p. 240-2.

159. Long H, Sung W, Kucukyildirim S, Williams E, Miller SF, Guo W, et al. Evolutionary determinants of genome-wide nucleotide composition. Nat Ecol Evol. 2018;2(2):237-40. doi: 10.1038/s41559-017-0425-y.
160. Desvaux M, Dalmasso G, Beyrouthy R, Barnich N, Delmas J, Bonnet R. Pathogenicity Factors of Genomic Islands in Intestinal and Extraintestinal Escherichia coli. Front Microbiol. 2020;11:2065. doi: 10.3389/fmicb.2020.02065.
161. Hacker J, Knapp S, Goebel W. Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an Escherichia coli O6 strain. J Bacteriol. 1983;154(3):1145-52. doi: 10.1128/jb.154.3.1145-1152.1983.
162. Brzuszkiewicz E, Brüggemann H, Liesegang H, Emmerth M, Olschläger T, Nagy G, et al. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic Escherichia coli strains. Proc Natl Acad Sci U S A. 2006;103(34):12879-84. doi: 10.1073/pnas.0603038103.
163. Lloyd AL, Henderson TA, Vigil PD, Mobley HL. Genomic islands of uropathogenic Escherichia coli contribute to virulence. J Bacteriol. 2009;191(11):3469-81. doi: 10.1128/jb.01717-08.
164. Tourret J, Diard M, Garry L, Matic I, Denamur E. Effects of single and multiple pathogenicity island deletions on uropathogenic Escherichia coli strain 536 intrinsic extra-intestinal virulence. Int J Med Microbiol. 2010;300(7):435-9. doi: 10.1016/j.ijmm.2010.04.013.
165. Schubert S, Rakin A, Heesemann J. The Yersinia high-pathogenicity island (HPI): evolutionary and functional aspects. Int J Med Microbiol. 2004;294(2-3):83-94. doi: 10.1016/j.ijmm.2004.06.026.
166. Sabaté M, Moreno E, Pérez T, Andreu A, Prats G. Pathogenicity island markers in commensal and uropathogenic Escherichia coli isolates. Clin Microbiol Infect. 2006;12(9):880-6. doi: 10.1111/j.1469-0691.2006.01461.x.
167. Buchholz U, Bernard H, Werber D, Böhmer MM, Remschmidt C, Wilking H, et al. German outbreak of Escherichia coli O104:H4 associated with sprouts. N Engl J Med. 2011;365(19):1763-70. doi: 10.1056/NEJMoa1106482.
168. Santos ACM, Santos FF, Silva RM, Gomes TAT. Diversity of Hybrid- and Hetero-Pathogenic Escherichia coli and Their Potential Implication in More Severe Diseases. Front Cell Infect Microbiol. 2020;10:339. doi: 10.3389/fcimb.2020.00339.
169. Outbreak of Escherichia coli O104:H4 infections associated with sprout consumption - Europe and North America, May-July 2011. MMWR Morb Mortal Wkly Rep. 2013;62(50):1029-31.
170. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, et al. Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. N Engl J Med. 2011;365(19):1771-80. doi: 10.1056/NEJMoa1106483.
171. Navarro-Garcia F. Escherichia coli O104:H4 Pathogenesis: an Enteroaggregative E. coli/Shiga Toxin-Producing E. coli Explosive Cocktail of High Virulence. Microbiol Spectr. 2014;2(6). doi: 10.1128/microbiolspec.EHEC-0008-2013.
172. Dutta S, Pazhani GP, Nataro JP, Ramamurthy T. Heterogenic virulence in a diarrheagenic Escherichia coli: evidence for an EPEC expressing heat-labile toxin of ETEC. Int J Med Microbiol. 2015;305(1):47-54. doi: 10.1016/j.ijmm.2014.10.006.
173. Mariani-Kurkdjian P, Lemaître C, Bidet P, Perez D, Boggini L, Kwon T, et al. Haemolytic-uraemic syndrome with bacteraemia caused by a new hybrid Escherichia coli pathotype. New Microbes New Infect. 2014;2(4):127-31. doi: 10.1002/nmi2.49.
174. World Health O. Antimicrobial resistance: global report on surveillance. Geneva: World Health Organization; 2014 2014.
175. Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. Microbiol Spectr. 2016;4(2). doi: 10.1128/microbiolspec.VMBF-0016-2015.

176. Clatworthy AE, Pierson E, Hung DT. Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol. 2007;3(9):541-8. doi: 10.1038/nchembio.2007.24.
177. Silver LL. Challenges of antibacterial discovery. Clin Microbiol Rev. 2011;24(1):71-109. doi: 10.1128/cmr.00030-10.
178. WHO. New report calls for urgent action to avert antimicrobial resistance crisis [13 July 2022]. Available from: https://www.who.int/news-room/detail/29-04-2019-new-report-calls-for-urgent-action-to-avert-antimicrobial-resistance-crisis.
179. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations 2016 [Available from: https://apo.org.au/node/63983.
180. Mathers AJ, Peirano G, Pitout JD. Escherichia coli ST131: The quintessential example of an international multiresistant high-risk clone. Adv Appl Microbiol. 2015;90:109-54. doi: 10.1016/bs.aambs.2014.09.002.
181. Codjoe FS, Donkor ES. Carbapenem Resistance: A Review. Med Sci (Basel). 2017;6(1). doi: 10.3390/medsci6010001.
182. Gao R, Hu Y, Li Z, Sun J, Wang Q, Lin J, et al. Dissemination and Mechanism for the MCR-1 Colistin Resistance. PLoS Pathog. 2016;12(11):e1005957. doi: 10.1371/journal.ppat.1005957.
183. McGann P, Snesrud E, Maybank R, Corey B, Ong AC, Clifford R, et al. Escherichia coli Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First Report of mcr-1 in the United States. Antimicrob Agents Chemother. 2016;60(7):4420-1. doi: 10.1128/aac.01103-16.
184. Schwarz S, Johnson AP. Transferable resistance to colistin: a new but old threat. J Antimicrob Chemother. 2016;71(8):2066-70. doi: 10.1093/jac/dkw274.
185. Kang CI, Kim J, Park DW, Kim BN, Ha US, Lee SJ, et al. Clinical Practice Guidelines for the Antibiotic Treatment of Community-Acquired Urinary Tract Infections. Infect Chemother. 2018;50(1):67-100. doi: 10.3947/ic.2018.50.1.67.
186. Organisation WH. WHO publishes list of bacteria for which new antibiotics are urgently needed 2017 [Available from: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.
187. Bush K. Proliferation and significance of clinically relevant β-lactamases. Ann N Y Acad Sci. 2013;1277:84-90. doi: 10.1111/nyas.12023.
188. Fleming A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. Br J Exp Pathol. 1929;10(3):226-36.
189. Bush K. Past and Present Perspectives on β-Lactamases. Antimicrob Agents Chemother. 2018;62(10). doi: 10.1128/aac.01076-18.
190. Mathers AJ, Peirano G, Pitout JD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clin Microbiol Rev. 2015;28(3):565-91. doi: 10.1128/cmr.00116-14.
191. Nicolas-Chanoine MH, Bertrand X, Madec JY. Escherichia coli ST131, an Intriguing Clonal Group. Clinical Microbiology Reviews. 2014;27(3):543-74. doi: 10.1128/cmr.00125-13.
192. Roer L, Overballe-Petersen S, Hansen F, Schønning K, Wang M, Røder BL, et al. Escherichia coli Sequence Type 410 Is Causing New International High-Risk Clones. mSphere. 2018;3(4). doi: 10.1128/mSphere.00337-18.
193. Hossain M, Tabassum T, Rahman A, Hossain A, Afroze T, Momen AMI, et al. Genotype-phenotype correlation of β-lactamase-producing uropathogenic Escherichia coli (UPEC) strains from Bangladesh. Sci Rep. 2020;10(1):14549. doi: 10.1038/s41598-020-71213-5.

194. Bartoletti R, Cai T, Wagenlehner FM, Naber K, Bjerklund Johansen TE. Treatment of Urinary Tract Infections and Antibiotic Stewardship. European Urology Supplements. 2016;15(4):81-7. doi: 10.1016/j.eursup.2016.04.003.
195. Kot B. Antibiotic Resistance Among Uropathogenic Escherichia coli. Pol J Microbiol. 2019;68(4):403-15. doi: 10.33073/pjm-2019-048.
196. Huovinen P, Sundström L, Swedberg G, Sköld O. Trimethoprim and sulfonamide resistance. Antimicrob Agents Chemother. 1995;39(2):279-89. doi: 10.1128/aac.39.2.279.
197. Sköld O. Resistance to trimethoprim and sulfonamides. Vet Res. 2001;32(3-4):261-73. doi: 10.1051/vetres:2001123.
198. Drozdowska D. 7.16 - Trimethoprim and Its Derivatives. In: Kenakin T, editor. Comprehensive Pharmacology. Oxford: Elsevier; 2022. p. 271-94.
199. Blaettler L, Mertz D, Frei R, Elzi L, Widmer AF, Battegay M, et al. Secular trend and risk factors for antimicrobial resistance in Escherichia coli isolates in Switzerland 1997-2007. Infection. 2009;37(6):534-9. doi: 10.1007/s15010-009-8457-0.
200. Critchley IA, Cotroneo N, Pucci MJ, Mendes R. The burden of antimicrobial resistance among urinary tract isolates of Escherichia coli in the United States in 2017. PLoS One. 2019;14(12):e0220265. doi: 10.1371/journal.pone.0220265.
201. Watts V, Brown B, Ahmed M, Charlett A, Chew-Graham C, Cleary P, et al. Routine laboratory surveillance of antimicrobial resistance in community-acquired urinary tract infections adequately informs prescribing policy in England. JAC Antimicrob Resist. 2020;2(2):dlaa022. doi: 10.1093/jacamr/dlaa022.
202. AURA 2021: Fourth Australian report on antimicrobial use and resistance in human health: Australian Commission on Safety and Quality in Health Care (ACSQHC); 2021 [Available from: https://www.safetyandquality.gov.au/sites/default/files/2021-09/aura_2021_-_report_-_final_accessible_pdf_-_for_web_publication.pdf.
203. Parry CM, Threlfall EJ. Antimicrobial resistance in typhoidal and nontyphoidal salmonellae. Curr Opin Infect Dis. 2008;21(5):531-8. doi: 10.1097/QCO.0b013e32830f453a.
204. Camins BC, Marschall J, DeVader SR, Maker DE, Hoffman MW, Fraser VJ. The clinical impact of fluoroquinolone resistance in patients with E coli bacteremia. J Hosp Med. 2011;6(6):344-9. doi: 10.1002/jhm.877.
205. Blondeau JM. Fluoroquinolones: mechanism of action, classification, and development of resistance. Surv Ophthalmol. 2004;49 Suppl 2:S73-8. doi: 10.1016/j.survophthal.2004.01.005.
206. Zechiedrich EL, Cozzarelli NR. Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli. Genes Dev. 1995;9(22):2859-69. doi: 10.1101/gad.9.22.2859.
207. Blondeau JM. Expanded activity and utility of the new fluoroquinolones: a review. Clin Ther. 1999;21(1):3-40; discussion 1-2. doi: 10.1016/s0149-2918(00)88266-1.
208. Wang JC, Lynch AS. Transcription and DNA supercoiling. Curr Opin Genet Dev. 1993;3(5):764-8. doi: 10.1016/s0959-437x(05)80096-6.
209. Kato J, Suzuki H, Ikeda H. Purification and characterization of DNA topoisomerase IV in Escherichia coli. J Biol Chem. 1992;267(36):25676-84.
210. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, et al. Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. J Infect Dis. 2013;207(6):919-28. doi: 10.1093/infdis/jis933.
211. Xiong X, Bromley EH, Oelschlaeger P, Woolfson DN, Spencer J. Structural insights into quinolone antibiotic resistance mediated by pentapeptide repeat proteins: conserved surface loops direct the activity of a Qnr protein from a gram-negative bacterium. Nucleic Acids Res. 2011;39(9):3917-27. doi: 10.1093/nar/gkq1296.

212. Browne AJ, Chipeta MG, Haines-Woodhouse G, Kumaran EPA, Hamadani BHK, Zaraa S, et al. Global antibiotic consumption and usage in humans, 2000-18: a spatial modelling study. Lancet Planet Health. 2021;5(12):e893-e904. doi: 10.1016/s2542-5196(21)00280-1.
213. Li X, Fan H, Zi H, Hu H, Li B, Huang J, et al. Global and Regional Burden of Bacterial Antimicrobial Resistance in Urinary Tract Infections in 2019. J Clin Med. 2022;11(10). doi: 10.3390/jcm11102817.
214. Douarre PE, Mallet L, Radomski N, Felten A, Mistou MY. Analysis of COMPASS, a New Comprehensive Plasmid Database Revealed Prevalence of Multireplicon and Extensive Diversity of IncF Plasmids. Front Microbiol. 2020;11:483. doi: 10.3389/fmicb.2020.00483.
215. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother. 2018;73(5):1121-37. doi: 10.1093/jac/dkx488.
216. Yang QE, Sun J, Li L, Deng H, Liu BT, Fang LX, et al. IncF plasmid diversity in multi-drug resistant Escherichia coli strains from animals in China. Front Microbiol. 2015;6:964. doi: 10.3389/fmicb.2015.00964.
217. Cohen SN. Transposable genetic elements and plasmid evolution. Nature. 1976;263(5580):731-8. doi: 10.1038/263731a0.
218. Garcillán-Barcia MP, Alvarado A, de la Cruz F. Identification of bacterial plasmids based on mobility and plasmid population biology. FEMS Microbiol Rev. 2011;35(5):936-56. doi: 10.1111/j.1574-6976.2011.00291.x.
219. Fortini D, Fashae K, García-Fernández A, Villa L, Carattoli A. Plasmid-mediated quinolone resistance and β-lactamases in Escherichia coli from healthy animals from Nigeria. J Antimicrob Chemother. 2011;66(6):1269-72. doi: 10.1093/jac/dkr085.
220. Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J, Voets GM, van den Munckhof MP, van Essen-Zandbergen A, et al. Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. Clin Microbiol Infect. 2011;17(6):873-80. doi: 10.1111/j.1469-0691.2011.03497.x.
221. Yang YQ, Li YX, Song T, Yang YX, Jiang W, Zhang AY, et al. Colistin Resistance Gene mcr-1 and Its Variant in Escherichia coli Isolates from Chickens in China. Antimicrob Agents Chemother. 2017;61(5). doi: 10.1128/aac.01204-16.
222. Yin W, Li H, Shen Y, Liu Z, Wang S, Shen Z, et al. Novel Plasmid-Mediated Colistin Resistance Gene mcr-3 in Escherichia coli. mBio. 2017;8(3). doi: 10.1128/mBio.00543-17.
223. Falgenhauer L, Waezsada SE, Yao Y, Imirzalioglu C, Käsbohrer A, Roesler U, et al. Colistin resistance gene mcr-1 in extended-spectrum β-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. Lancet Infect Dis. 2016;16(3):282-3. doi: 10.1016/s1473-3099(16)00009-8.
224. Toleman MA, Walsh TR. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. FEMS Microbiol Rev. 2011;35(5):912-35. doi: 10.1111/j.1574-6976.2011.00294.x.
225. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38(5):865-91. doi: 10.1111/1574-6976.12067.
226. Stokes HW, Gillings MR. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. FEMS Microbiol Rev. 2011;35(5):790-819. doi: 10.1111/j.1574-6976.2011.00273.x.
227. Varani A, He S, Siguier P, Ross K, Chandler M. The IS6 family, a clinically important group of insertion sequences including IS26. Mob DNA. 2021;12(1):11. doi: 10.1186/s13100-021-00239-x.

228. Hagemann AT, Craig NL. Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site. Genetics. 1993;133(1):9-16. doi: 10.1093/genetics/133.1.9.
229. Iyer A, Barbour E, Azhar E, Salabi A, Hassan H, Qadri I, et al. Transposable elements in Escherichia coli antimicrobial resistance. Advances in Bioscience and Biotechnology. 2013;4:415-23.
230. Babakhani S, Oloomi M. Transposons: the agents of antibiotic resistance in bacteria. J Basic Microbiol. 2018;58(11):905-17. doi: 10.1002/jobm.201800204.
231. 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). doi: 10.1128/mBio.02381-17.
232. Bednorz C, Oelgeschläger K, Kinnemann B, Hartmann S, Neumann K, Pieper R, et al. The broader context of antibiotic resistance: zinc feed supplementation of piglets increases the proportion of multi-resistant Escherichia coli in vivo. Int J Med Microbiol. 2013;303(6-7):396-403. doi: 10.1016/j.ijmm.2013.06.004.
233. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. Microbiol Mol Biol Rev. 1999;63(3):507-22. doi: 10.1128/mmbr.63.3.507-522.1999.
234. Lévesque C, Piché L, Larose C, Roy PH. PCR mapping of integrons reveals several novel combinations of resistance genes. Antimicrob Agents Chemother. 1995;39(1):185-91. doi: 10.1128/aac.39.1.185.
235. Halaji M, Feizi A, Mirzaei A, Sedigh Ebrahim-Saraie H, Fayyazi A, Ashraf A, et al. The Global Prevalence of Class 1 Integron and Associated Antibiotic Resistance in Escherichia coli from Patients with Urinary Tract Infections, a Systematic Review and Meta-Analysis. Microb Drug Resist. 2020;26(10):1208-18. doi: 10.1089/mdr.2019.0467.
236. Gillings MR. Integrons: Past, Present, and Future. Microbiology and Molecular Biology Reviews. 2014;78(2):257-77. doi: 10.1128/mmbr.00056-13.
237. Stokes HW, Hall RM. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Mol Microbiol. 1989;3(12):1669-83. doi: 10.1111/j.1365-2958.1989.tb00153.x.
238. Carattoli A. Importance of integrons in the diffusion of resistance. Vet Res. 2001;32(3-4):243-59. doi: 10.1051/vetres:2001122.
239. 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. Microb Genom. 2017;3(12). doi: 10.1099/mgen.0.000143.
240. Harmer CJ, Hall RM. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. mSphere. 2016;1(2). doi: 10.1128/mSphere.00038-16.
241. Cambray G, Guerout AM, Mazel D. Integrons. Annu Rev Genet. 2010;44:141-66. doi: 10.1146/annurev-genet-102209-163504.
242. Mazel D. Integrons: agents of bacterial evolution. Nat Rev Microbiol. 2006;4(8):608-20. doi: 10.1038/nrmicro1462.
243. 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). doi: 10.1099/mgen.0.000255.
244. Meek RW, Vyas H, Piddock LJ. Nonmedical Uses of Antibiotics: Time to Restrict Their Use? PLoS Biol. 2015;13(10):e1002266. doi: 10.1371/journal.pbio.1002266.
245. Wales AD, Davies RH. Co-Selection of Resistance to Antibiotics, Biocides and Heavy Metals, and Its Relevance to Foodborne Pathogens. Antibiotics (Basel). 2015;4(4):567-604. doi: 10.3390/antibiotics4040567.
246. Kemper N. Veterinary antibiotics in the aquatic and terrestrial environment. Ecological Indicators. 2008;8(1):1-13. doi: https://doi.org/10.1016/j.ecolind.2007.06.002.

247. Reinthaler FF, Posch J, Feierl G, Wüst G, Haas D, Ruckenbauer G, et al. Antibiotic resistance of E. coli in sewage and sludge. Water Res. 2003;37(8):1685-90. doi: 10.1016/s0043-1354(02)00569-9.
248. Kraemer SA, Ramachandran A, Perron GG. Antibiotic Pollution in the Environment: From Microbial Ecology to Public Policy. Microorganisms. 2019;7(6). doi: 10.3390/microorganisms7060180.
249. Wang SH, Sheng WH, Chang YY, Wang LH, Lin HC, Chen ML, et al. Healthcare-associated outbreak due to pan-drug resistant Acinetobacter baumannii in a surgical intensive care unit. J Hosp Infect. 2003;53(2):97-102. doi: 10.1053/jhin.2002.1348.
250. Su S, Li C, Yang J, Xu Q, Qiu Z, Xue B, et al. Distribution of Antibiotic Resistance Genes in Three Different Natural Water Bodies-A Lake, River and Sea. Int J Environ Res Public Health. 2020;17(2). doi: 10.3390/ijerph17020552.
251. Zurek L, Ghosh A. Insects represent a link between food animal farms and the urban environment for antibiotic resistance traits. Appl Environ Microbiol. 2014;80(12):3562-7. doi: 10.1128/aem.00600-14.
252. Tyrrell C, Burgess CM, Brennan FP, Walsh F. Antibiotic resistance in grass and soil. Biochem Soc Trans. 2019;47(1):477-86. doi: 10.1042/bst20180552.
253. Laborda P, Sanz-García F, Ochoa-Sánchez LE, Gil-Gil T, Hernando-Amado S, Martínez JL. Wildlife and Antibiotic Resistance. Front Cell Infect Microbiol. 2022;12:873989. doi: 10.3389/fcimb.2022.873989.
254. Robinson TP, Bu DP, Carrique-Mas J, Fèvre EM, Gilbert M, Grace D, et al. Antibiotic resistance is the quintessential One Health issue. Trans R Soc Trop Med Hyg. 2016;110(7):377-80. doi: 10.1093/trstmh/trw048.
255. Hernando-Amado S, Coque TM, Baquero F, Martínez JL. Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nat Microbiol. 2019;4(9):1432-42. doi: 10.1038/s41564-019-0503-9.
256. McMichael C. Climate change-related migration and infectious disease. Virulence. 2015;6(6):548-53. doi: 10.1080/21505594.2015.1021539.
257. Anjum MF, Schmitt H, Börjesson S, Berendonk TU. The potential of using E. coli as an indicator for the surveillance of antimicrobial resistance (AMR) in the environment. Curr Opin Microbiol. 2021;64:152-8. doi: 10.1016/j.mib.2021.09.011.

Chapter 3: Overview of the methodology

3.1. Study design

Please be aware that the information provided in this chapter describes a general workflow performed by the candidate. Each results chapter contains a detailed "materials and methods" section relevant to it.

Genomic epidemiology is a discipline which uses pathogen genomic data to determine the distribution and spread of an infectious disease in a particular population and uses this information to assist in the management of health problems. Genomic epidemiology allows researchers to track and study the dissemination of AMR and associated genomic elements in microbes, including *E. coli,* across human, animal, and environmental populations. Genomic epidemiology provides much higher resolution and at the same time higher reproducibility and scalability when compared to traditional methods like PCR (1). WGS, and in particular short-read sequencing (like Illumina platforms), provide a cost effective, accurate and high-throughput means to produce the data required for genomic epidemiological studies (2).

3.2. Collection overview

The primary collection was comprised of 451 *E. coli* isolates sourced from urine specimens submitted for culturing to the Microbiology department of Orange Base Hospital (OBH), NSW, Australia. These isolates were collected between 2006 to 2011, with most isolates originating from 2006 to 2009. This collection was submitted for WGS (see section 3.3) and post quality control measures (see section 3.4.1), the study collection totalled 426 whole genome sequences. Of these 426 isolates, 67 were phenotypically tested for trimethoprim resistance and studied in Chapter 5, 23 isolates were previously identified as ST131 by PCR and studied in Chapter 6, while the remaining 336 isolates were studied in Chapter 4.

Medical personnel at the participating healthcare centres collected urine sample following a standardised protocol. Semi-quantitative urine culture was performed on horse blood, MacConkey and chromogenic agars. *E. coli* were identified by conventional biochemical tests. After that isolates were stored in 50% (v/v) glycerol in trypticase soy broth at -70 °C.

3.3. DNA extraction and whole genome sequencing

Short-read WGS using Illumina requires small quantities of DNA (~1 ng), which are readily achieved using commercial DNA extraction kits. Glycerol stocks were used to inoculate sterile Lysogeny broth (LB) broth (~5 ml in 15 ml falcon tube) and incubated at 37^{o}C overnight. Approximately 1 ml of cell culture was then pelleted at 8000 x *g* for 5 min, lysed and DNA was

extracted using a Bioline™ ISOLATE II Genomic DNA Kit following manufacturer's standard protocol for bacterial cells, with modification of last step where DNase and RNase free water was used for DNA elution.

Most of the collection was sequenced using an Illumina NovaSeq. Illumina sequencing technology is the gold standard of short sequencing, providing highly accurate base calls. However, this sequencing technique has the drawback of providing reads of short length, which are problematic in assembling CRRs as these tend to contain many repetitive sequences which cause contig breaks. Nevertheless, short-read sequencing remains far more cost-effect compared to long-read sequencing (3), especially when hundreds of genomes are involved.

3.4. Bioinformatical methods

Computational analyses were conducted using the Interactive High-Performance Computing facility at UTS, which utilises a virtual machine-based setup for node control, rather than traditional scheduling systems. All computations performed during candidature could be done using consumer-grade hardware, with the primary limitations being random access memory (RAM), storage capacity, and computational time. The code and commands used in this study require no modifications for UNIX-based operating systems (Linux/Mac).

3.4.1. Genome quality control, assembly and annotation

To assemble all short-read sequence data, the Shovill v1.0.4 (github.com/tseemann/shovill) pipeline was used, with SPAdes (4) as the underlining assembler. The use of Shovill was justified by its high performance and ease of use, and the inclusion of features such as trimming and error correction (3).

Prior to assembly, sequence read quality was assessed using FastQC v0.11.9 (bioinformatics.babraham.ac.uk/projects/fastqc/) and then the quality of each assembly was assessed using assembly-stats software (github.com/sanger-pathogens/assembly-stats). Each genome had to meet three main criteria: a genome between 4.5 Mb - 6.5 Mb, an N50 value of more than 30 000 and have sequencing depth of 10X or more. The first criteria prevented the inclusion of contaminated genomes, as the expected *E. coli* genome size ranges from 4 to 6 Mb (5, 6), the second criterion, defined by the size of the smallest contig for which longer or equal size contigs cover ≥50% of the assembly, ensured relatively high contiguity, while third make sure that nucleotides are sequenced with required accuracy.

Rapid prokaryotic genome annotation pipeline Prokka (7) was then used for automated annotation of genome assemblies. Prokka is a fast and high throughput software with the

advantage of being able to be run on a local machine or a high-performance computational cluster. Based on performance benchmarks, Prokka provides better and faster results compared to other annotation software like RAST or xBASE2 (7).

3.4.2. Additional genomes from publicly available databases

Genomic comparative analyses were performed to provide insights into the greater genomic landscape encompassing the study collection. Additional genomes, including Australian ST131 isolates used in Chapter 6, and Australian ST73, ST95 and ST127 used in Chapter 4, were downloaded from publicly available sources, such as Enterobase (enterobase.warwick.ac.uk) (8), an integrated software, which includes a database containing more than 200 000 *Escherichia*/*Shigella* genomes. Enterobase is a web portal with assemblies and metadata repositories and contains additional analytical tools including MLST search and phylogenetic tree building software GrapeTree (9).

3.4.3. Populational structure

Multiple methods were used to establish phylogenetic relationships between isolates. The workflow to produce high resolution phylogenetic tree from short-read data consisted of three main steps: 1) produce draft genome assemblies (section 3.4.1); 2) produce multisequence alignments (sections 3.4.3.1 and 3.4.3.2) and 3) build phylogenetic trees using multisequence alignments (section 3.4.3.3). To determine all groups of orthologous genes within a population, as well as all accessory genes, pangenome analyses were performed (section 3.4.3.4).

3.4.3.1. Multi-locus sequence typing (MLST)

As mentioned in the introduction, MLST is a classification system used for bacteria, including *E. coli*. Multiple MLST schemes exist, however in this body of work, the Achtman scheme was employed (10, 11). MLST were assigned using mlst software (github.com/tseemann/mlst) with automatic scheme detection and default settings (12).

3.4.3.2. Single nucleotide polymorphism (SNP) based alignment

SNP based core genome phylogenies are excellent tools for reconstructing evolutionary relationships between closely related sequences such as strains of the same ST or CC. In this thesis, Snippy (github.com/tseemann/snippy) was used as it contains reliable and well-established tools such as Burrows-Wheeler Aligner (BWA) (13) and Freebayes (github.com/ekg/freebayes). Snippy is fast and produces a core genome alignment of strains to a reference genome. However, it is well known that the genomes of *E. coli* are often subjected to recombination, which introduces multiple SNPs in one genomic event, rendering closely

related strain further apart on a phylogenetic tree generated by this method. Therefore, it is important to utilise recombination filtering software, such as Gubbins, used in this thesis. Gubbins identifies regions with unexpectedly high level of SNPs, which serve as a proxy for recombination events, compares them to background levels of SNP accumulation, and removes them if necessary (14). The downside of this method is the use of a reference genome (usually well established and well annotated genome like *E. coli* K-12 MG1655 (GCA_000005845.2), which can produce strong biases. All genomes are aligned to the reference, which may not hold genetic elements representative for the collection studied (15).

3.4.3.3. Core genome-based alignment

Core genome phylogeny is an alternative method to SNP-based reference phylogenies. Core genome phylogenies generate reference-independent alignments and rely on automated genome annotation pipeline, in our case Prokka (7), with core genome alignment produced by Roary (16). The core genome is considered the set of genes present in all members of collection studied. By using a common denominator (i.e., the core genome), this method is absent of reference-based bias discussed previously. However, core genome-based alignments do depend on a genome annotation step, which can introduce errors, and ignore intergenic regions.

3.4.3.4. Phylogenetic tree generation and visualisation

Phylogenetic trees from either reference-based SNP or core genome alignments were generated using IQtree2 (17), with extended model selection and tree construction by best-fit model (-m MFP) and 1000 boot strap replicates (-bb 1000). IQtree2 is an open-sourced software package with high efficiency and performance (17, 18). The best-fit model was determined using ModelFinder, which is incorporated in IQtree2 and screens through multiple substitution models and determines which model best explains the evolution within data set using statistical criteria like Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (19). Visualisation of phylogenetic trees were done in the Interactive Tree of Life (iTOL) (20), a web-based service for display, annotation and management of phylogenetic trees.

3.4.3.5. Pan genome analysis

Pan genome analyses break down a collection of genomes into compartments: a core genome (usually defined as genes present in 99% of genomes), soft core genome (<99% and >=95% of genome), shell genome (<95% and >=15% of genomes) and cloud genome which contains all singlets (<15% of genomes). In this thesis, Roary software, using GFF3 files produced by Prokka, was used to extract and cluster similar genes in these four categories, in convenient csv format (16). Based on Roary outputs, the statistical software Scoary was used to quickly conduct GWA

studies, using any phenotypic trait known for isolates. Scoary provides statistically tested and p-value corrected for multiple comparison results for every phenotypic trait used (21).

3.4.4. Gene screening

By utilising WGS data, researchers can, quickly and *en masse,* identify the presence or absence of any genes of interest, including ARGs, VAGs, plasmid replicons and ISs. There are two methods to do that, assembly-based and short-read based.

Assembly-based method require the generation of a *de novo* assembly before gene identification can be performed. The majority of tools (e.g., Abricate (github.com/tseemann/abricate) utilise the National Centre for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (22), but can be used directly via an online portal or through a separate command line tool. The latter is useful as it enables large scale gene screening of entire collections against extensive gene databases. The major advantage of this approach is a speed and ability to run in parallel even on low end modern computers. However, the disadvantage is it depend on an assembler, leading to assembly bias, where errors or problems during assembly may result in a gene incorrectly determined or even not detected at all.

Short read-based approach aligning short reads to the reference genes, this method avoids assembly bias. As such, ARIBA (23) software was used in Chapter 4. Like Abricate, ARIBA can also perform large scale gene screening, however it is much slower compared to assembly-based methods.

Gene screening requires well curated databases of genes to be screened against. The Centre of Genomic Epidemiology (genomicepidemiology.org/) hosts several commonly used databases including ResFinder (24, 25), VirulenceFinder (26, 27), PlasmidFinder (28), pMLST (28) and SerotypeFinder (28). These databases can be used via online screening tools available at on Centre of Genomic Epidemiology website or downloaded locally and used with ARIBA or Abricate. Other more comprehensive ARG and VAG databases include Comprehensive Antimicrobial Resistance Database (CARD) (29) and the Virulence Factor Database (VFDB) (30).

3.4.5. Mobile Genetic Elements (MGEs)

In this thesis we studied three main types of MGEs: plasmids, PAIs and integrons.

3.4.5.1. Plasmid analysis

While completed plasmids are rarely resolved using short-reads, it is possible to identify plasmid Inc groups and perform RST for some of the more common Inc types, such as IncF plasmids. This data is useful in preliminary identification of particular lineages of plasmids, such as ColV and

pUTI89 plasmids. Mapping of reference plasmids was done in three ways in this thesis, 1) with the use of BWA on short reads from each sample with later production of histogram of read depth to reference coordinates (31), or 2) using Abricate with data containing plasmid of interest to generate a simplified BLAST output and later process it with plasmidmapR script (github.com/maxlcummins/plasmidmapR) (32), or 3) comparisons of plasmid sequences against reference plasmids using CGView web-based software (33).

3.4.5.2. Pathogenicity-associated islands analysis (PAIs)

PAIs are distinct genomic elements present in many pathogenic bacterial lineages. PAIs are considered as a subclass of genomic islands which are acquired by HGT and can provide "quantum leaps" in microbial evolution (34). PAI comparisons were performed in a similar way to plasmid comparison (section 3.4.5.1), with PAIDB (paidb.re.kr) (35) used as a database to source reference sequences of PAIs studied. PAI alignments were achieved using Abricate with plasmidmapR (chapter 4) or BWA based read mapping (chapter 5).

3.4.5.3. Integron analysis

In order to study integron structures the following approach was utilised: first, genomes which contain full or partial forms of class 1 or 2 intergon integrases were identified using local instance of BLAST+ (36). The contigs which contained full or partial class 1 or 2 integron integrases were exported and annotations were manually curated in SnapGene (v4.1.9) (GSL Biotech, USA) with use of the online database Uniprot (37), and final visualisation were done using open source Krita (krita.org) software.

3.5. References

1. 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. Front Microbiol. 2016;7:843. doi: 10.3389/fmicb.2016.00843.
2. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol. 2018;122(1):e59. doi: 10.1002/cpmb.59.
3. Warnke-Sommer J, Ali H. Graph mining for next generation sequencing: leveraging the assembly graph for biological insights. BMC Genomics. 2016;17:340. doi: 10.1186/s12864-016-2678-2.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021.
5. Kurylo CM, Alexander N, Dass RA, Parks MM, Altman RA, Vincent CT, et al. Genome Sequence and Analysis of Escherichia coli MRE600, a Colicinogenic, Nonmotile Strain that Lacks RNase I and the Type I Methyltransferase, EcoKI. Genome Biol Evol. 2016;8(3):742-52. doi: 10.1093/gbe/evw008.

6. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res. 2012;40(Database issue):D115-22. doi: 10.1093/nar/gkr1044.
7. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9. doi: 10.1093/bioinformatics/btu153.
8. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 2020;30(1):138-52. doi: 10.1101/gr.251678.119.
9. Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 2018;28(9):1395-404. doi: 10.1101/gr.232397.117.
10. Tartof SY, Solberg OD, Manges AR, Riley LW. Analysis of a uropathogenic Escherichia coli clonal group by multilocus sequence typing. J Clin Microbiol. 2005;43(12):5860-4. doi: 10.1128/jcm.43.12.5860-5864.2005.
11. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol. 2006;60(5):1136-51. doi: 10.1111/j.1365-2958.2006.05172.x.
12. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595. doi: 10.1186/1471-2105-11-595.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324.
14. 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. 2015;43(3):e15-e. doi: 10.1093/nar/gku1196.
15. Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front Genet. 2015;6:235. doi: 10.3389/fgene.2015.00235.
16. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3. doi: 10.1093/bioinformatics/btv421.
17. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530-4. doi: 10.1093/molbev/msaa015.
18. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188-95. doi: 10.1093/molbev/mst024.
19. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9. doi: 10.1038/nmeth.4285.
20. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-w6. doi: 10.1093/nar/gkab301.
21. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016;17(1):238. doi: 10.1186/s13059-016-1108-8.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10. doi: 10.1016/s0022-2836(05)80360-2.
23. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3(10):e000131. doi: 10.1099/mgen.0.000131.

24. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75(12):3491-500. doi: 10.1093/jac/dkaa345.
25. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother. 2017;72(10):2764-8. doi: 10.1093/jac/dkx217.
26. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol. 2014;52(5):1501-10. doi: 10.1128/jcm.03617-13.
27. Malberg Tetzschner AM, Johnson JR, Johnston BD, Lund O, Scheutz F. In Silico Genotyping of Escherichia coli Isolates for Extraintestinal Virulence Genes by Use of Whole-Genome Sequencing Data. J Clin Microbiol. 2020;58(10). doi: 10.1128/jcm.01269-20.
28. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903. doi: 10.1128/aac.02412-14.
29. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. J Clin Microbiol. 2015;53(8):2410-26. doi: 10.1128/jcm.00008-15.
30. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2022;50(D1):D912-d7. doi: 10.1093/nar/gkab1107.
31. 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). doi: 10.1099/mgen.0.000250.
32. Cummins ML, Reid CJ, Djordjevic SP. F Plasmid Lineages in Escherichia coli ST95: Implications for Host Range, Antibiotic Resistance, and Zoonoses. mSystems. 2022;7(1):e0121221. doi: 10.1128/msystems.01212-21.
33. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. Bioinformatics. 2005;21(4):537-9. doi: 10.1093/bioinformatics/bti054.
34. Gal-Mor O, Finlay BB. Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell Microbiol. 2006;8(11):1707-19. doi: 10.1111/j.1462-5822.2006.00794.x.
35. Yoon SH, Park YK, Kim JF. PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res. 2015;43(Database issue):D624-30. doi: 10.1093/nar/gku985.
36. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi: 10.1186/1471-2105-10-421.
37. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480-d9. doi: 10.1093/nar/gkaa1100.

Chapter 4: Dominance of Escherichia coli sequence types ST73, ST95, ST127, and ST131 in urine isolates: a genomic analysis of antimicrobial resistance and virulence linked to F plasmids

4.1. Declaration

This chapter represents a manuscript accepted for publication.

Dmitriy Li[1,2], Paarthiphan Elankumaran[1,2], Timothy Kudinha[3], Amanda K. Kidsley[4], Darren J. Trott[4], Veronica M. Jarocki[1,2*], Steven P. Djordjevic[1,2*]

[1] Australian Institute for Microbiology & Infection, University of Technology Sydney, Ultimo, NSW, Australia
[2] Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, NSW, Australia
[3] Central West Pathology Laboratory, Charles Sturt University, Orange, NSW, Australia
[4] School of Animal and Veterinary Science, The University of Adelaide, Adelaide, South Australia, Australia

Accepted for publication in:
Microbial Genomics, June 2023

Autor contributions:
D.L.: analysis and interpretation of data, drafted and revised the work; P.E.: acquisition of data; T.K.: acquisition of data; A.K.K.: acquisition of data and revised the work; D.J.T.: conception and design of work, acquisition of data and revised the work; V.M.J.: conception and design of work, interpretation of data, supervision, drafted and revised the work; S.P.D.: conception and design of work, acquisition of data, supervision, drafted and revised the work.

Ethics approval and consent to participate:
The project was approved by Charles Sturt University and Sydney West Area Health Service research ethics committees. Since clinical information for patients with UTIs was provided anonymously by clinicians, patient consent was not required.

4.2. Abstract

Extraintestinal *Escherichia coli* is the most frequent cause of UTIs globally. Most studies of clinical *E. coli* isolates are selected on AMR phenotypes; however, this selection bias may not provide an accurate portrayal of which STs cause the most disease. Here, WGS was performed on 336 *E. coli* isolates from urine samples sourced from a regional hospital in Australia in 2006. Most isolates (91%) were sourced from patients with UTIs and were not selected based on any AMR phenotypes. No significant differences were observed in AMR profiles across age, sex and uro-clinical syndromes, though isolates derived from children (0-11 yrs) carried significantly more (p=0.016) VAGs. While 91 ST were identified, ST73, ST95, ST127 and ST131 dominated. F virulence plasmids carrying *senB-cjrABC* virulence genes were a feature of this collection. These *senB-cjrABC*+ plasmids were split into two categories: pUTI89-like (F29:A-:B10 and/or > 95%

identity to pUTI89) (n=78) and non-pUTI89-like (n=53). Compared to all other plasmid replicons, isolates with pUTI89-like plasmids carried fewer ARGs, whilst isolates with *senB-cjrABC*+/non-pUTI89 plasmids had a significantly higher load of ARGs and class 1 integrons. F plasmids were not detected in 95 genomes, predominantly ST73. Our phylogenomic analyses identified closely related isolates from the same patient associated with different pathologies.

4.3. Impact Statement

Pandemic ExPEC lineages, such as ST ST131, have been largely responsible for a global increase in ESBL-producing *E. coli*, which are a WHO critical priority pathogen. Due to the importance of AMR to both individual patient care and public health, most studies of clinical *E. coli* isolates are selected on AMR phenotypes. However, this selective sampling creates selection bias and skews baseline data which could otherwise be used to predict AMR trends and trajectories. This retrospective study implemented detailed phylogenetics, SNP and pangenome analyses, gene screening and plasmid and pathogenicity island mapping on one of the largest cohorts of *E. coli* from clinical urine samples without selection biases and highlights the dominance of *E. coli* carrying F plasmids with *senB-cjrABC* virulence genes in UTIs.

4.4. Introduction

UTIs incur an enormous cost burden to society and are the leading clinical presentation that drives antibiotic prescription (1). In Australia, UTIs cause an estimated 69,823 annual hospitalisations (2) and cost the nation's health system AU$909 million annually (3). ExPEC are the leading cause of UTIs and are the most frequently isolated Gram-negative pathogen globally (4). Additionally, ExPEC are responsible for bloodstream and wound infections, neonatal meningitis, and is the most frequent cause of ventilator-associated pneumonia (5). Of more than 13,000 *E. coli* STs, only a small subset of pandemic lineages are responsible for the vast majority of ExPEC infections (6). Pandemic ExPEC lineages, such as ST131, have also been central to a global increase in ESBL-producing *E. coli*, as well as resistance to other clinically important antibiotic classes (4, 7, 8). A combination of sulphonamide and trimethoprim is a standard treatment for UTIs but resistance rates to these frontline antibiotics are increasing globally leading to the elevated use of extended-spectrum β-lactams and fluoroquinolones (9-13).

The gut is a major reservoir of ExPEC, with vast numbers of these organisms shed into wastewater and diverse agricultural environments where they become exposed to frequent and often constant antimicrobial selection pressures, particularly in municipal wastewater, food animal production, and in animal faecal holding ponds. Gastrointestinal carriage of major ExPEC clonal lineages is influenced by frequent host-to-host transmission facilitated by sexual contact,

international travel, contaminated food and water consumption, and interactions with wildlife, livestock, and companion animals (15-19). Constant recolonisation with and across different hosts and repeated exposure to aquatic and terrestrial environments undoubtedly influence how *E. coli* acquires genetic information by HGT. Understanding the genetic features that enable colonisation of different hosts, food, livestock, companion animals, wildlife, and water sanitation practices profoundly influences pathogen biology, particularly regarding the evolution of emerging lineages, and is at the forefront of predictive infectious disease management. Much of the value inherent in studying successful pandemic ExPEC lineages lies in understanding what genetic features can be attributed to their global success, notwithstanding biological success can evolve by diverse and often convergent paths. What is clear is that the acquisition of MGEs, particularly F plasmids (18, 20-22), and genomic islands (23, 24), underpins lineage evolution while AMR affords opportunities for niche persistence.

The association of F virulence plasmids, such as pUTI89 and ColV, in *E. coli* lineage evolution, host range, and zoonosis is an increasingly important area of enquiry (18). A detailed analysis linking the carriage of these plasmids in urinary tract isolates of *E. coli* has not been conducted. In pUTI89 and related plasmids, carriage of the *cjr* operon and the putative enterotoxin gene *senB* are considered important for virulence (25). A study that interrogated a cohort of 34,176 *E. coli* genome sequences (2,570 STs) showed pUTI89, (RST F29:A-:B10) was overwhelmingly linked to *E. coli* sourced from humans but was almost entirely absent from 13,027 *E. coli* isolates recovered from poultry, pigs, and cattle (18). F plasmids with RST F29:A-:B10 have been associated with specific sublineages of major pandemic ExPEC lineages ST131, ST73, ST69, and ST95 (14, 18, 20). In that same study, ColV-like plasmids were represented among *E. coli* sourced from poultry (2,327/4,254 [55%]) but also evident in 720/4,425 (16%) human ExPEC isolates (18). ColV virulence plasmids are found in *E. coli* that: i) cause extraintestinal disease in humans and poultry as well as commensal *E. coli* (18, 26-28); ii) are required for APEC to cause colibacillosis (29); iii) have been linked with zoonotic *E. coli* infections (18), and; iv) display resistance to chlorine (30). However, incomplete metadata linked with isolates deposited in public databases precludes a more thorough investigation of the association of these plasmids in *E. coli* linked to UTIs. This observation has clear implications in shedding light on a deeper understanding of the One Health aspects of *E. coli* disease.

Here we have undertaken a comprehensive phylogenomic analysis of 336 *E. coli* isolates from the urine of patients experiencing different clinical afflictions but predominantly with a uro-pathological focus (cystitis, pyelonephritis) from a single rural hospital in NSW, Australia.

4.5. Data summary

WGS of 336 isolates described here were uploaded to the Sequence Read Archive (SRA) in the NCBI, under BioProject number PRJNA842786 with BioSample accession numbers SAMN28687489 to SAMN28687824. WGS of 24 companion animal isolates were also uploaded to the SRA under BioProject number PRJNA858941. These 24 isolates were previously analysed in Saputra *et al*. (2017). Antimicrobial resistance in clinical *Escherichia coli* isolated from companion animals in Australia. Vet Microbiol.;211:43-50. doi: 10.1016/j.vetmic.2017.09.014

4.6. Methods

4.6.1. Sample collection

Urine samples from patients presenting at OBH were collected over a six-month period (13th May 2006 to 12th November 2006). Specimens were included if they yielded a bacterial count of > 108 cfu/L and cell count of > 108/L for white blood cells and < 103/L for epithelial cells. Specimens were excluded if patients had known diabetes mellitus, diarrhea, received antibiotic therapy in the last month prior admission, or were menstruating. Using this inclusion/exclusion criteria, a total of 353 samples (326 midstream urine [MSU]; 27 catheter specimen urine [CSU]) were collected from 322 patients. Urine specimens were collected as previously described (14). Briefly, each participating physician received a protocol for urine collection and the diagnostic criteria for classification of the uro-clinical syndrome. A diagnosis of cystitis or pyelonephritis required specific manifestations, as recorded by the treating medical practitioner. Cystitis defining manifestations included dysuria, frequent urination, and/or suprapubic tenderness, without fever or loin pain. Pyelonephritis defining manifestations included urinary symptoms with a fever of ≥ 38°C and flank pain, with or without nausea/vomiting. Semi-quantitative cultures were performed on horse blood, MacConkey, and chromogenic agars, followed by conventional biochemical tests. Isolates were stored in 50% (v/v) glycerol in trypticase soy broth at −70°C until further use.

4.6.2. Whole genome sequencing and genome assembly

DNA extraction and sequencing were performed as described previously (31). Briefly, DNA was extracted using an ISOLATE II Genomic DNA (Bioline) kit following the manufacturer's standard protocol for bacterial cells, except for the final DNA elution step in which DNase and RNase free water were used. Library preparation was performed by the UTS Core Sequencing Facility at the University of Technology Sydney, using the adapted Nextera Flex library preparation kit process, Hackflex (32). Sequencing was performed using Illumina Novaseq S4 flow cell, 2 × 150 bp at Novogene (Singapore). The quality of reads was assessed using fastp (v0.20.1) (33).

4.6.3. Phylogeny and SNP analyses

Maximum-likelihood core genome phylogenies were built using IQtree2 (v2.0.3) (34) with extended model selection and tree reconstruction by best-fit model (-m MFP), 1000 bootstrap replicates (-bb 1000), using a core genome alignment generated by Roary (v3.13.0) (35) and the fast core gene alignment with MAFFT (-e --mafft) option. All core genome phylogenies were confirmed using both marker gene-based phylogenies produced by Phylosift using the default settings (36), and SNP-based phylogenies produced by snippy v4.3.6 (github.com/tseemann/snippy) with default settings. Phylogroups were determined by EzClermont (37). SNP heatmaps were built by making pairwise SNP distance matrices of core genome alignments produced by Roary using snp-dists v0.6.3 with default settings (github.com/tseemann/snp-dists) and visualised using R v4.1.2.

4.6.4. Gene screening

All isolates were annotated using the prokka pipeline (38) set for *E. coli* (--genus Escherichia --species coli), forced Genbank/ENA/DDJB compliance (--compliant), standard bacterial genetic code (translation table 11) (--gcode 11). All isolates were screened for ARGs, VAGs, the presence of ColV markers, MLSTs (Achtman 7 Gene), and pMLST as previously described (18), using a pipeline available at github.com/maxlcummins/pipelord2_0. This pipeline incorporates assembly-stats v1.0.1 (github.com/sanger-pathogens/assembly-stats), fastp v0.20.1, kraken2 v2.2.1 (39), mlst v2.19.0 (github.com/tseemann/mlst) with default setting including automatic detection of best fit MLST scheme, and pointfinder (40). Assemblies with a genome size of < 4.5 Mbp and > 6.5 Mbp and N50 > 30000 were excluded. Gene screening was performed using Abricate v1.0.1 (github.com/tseemann/abricate) in conjunction with the following databases: VFDB (41), CARD (42), ISfinder (isfinder.biotoul.fr/), and PlasmidFinder (43). Genes were marked as present when detected by Abricate and filtered using a custom R script (github.com/maxlcummins/abricateR) to have 90% identity and 90% coverage. The presence and length of the class 1 integron integrase *intI1* was determined using BLAST+ v2.8.1 (44). Contigs harbouring class 1 integrons were further analysed in SnapGene v4.1.9 (snapgene.com) and an integron map was drawn using graphics software Krita v4.4.1 (krita.org). Regarding specific IncF plasmid type groupings - as per modified Liu *et al*., (2018) criteria (45), isolates were marked as being ColV plasmid positive if they contained at least one gene from four or more of the following sets (i) *cvaABC* and *cvi*, (ii) *iroBCDEN*, (iii) *iucABCD* and *iutA*, (iv) *etsABC*, (v) *ompT* and *hlyF*, and (vi) *sitABCD* (45), with ≥ 95% identity and ≥ 95% gene coverage. Isolates with either ≥ 90% length and identity of pUTI89 (NC_007941.1) or having IncF RST F29:A-:B10 were considered to be a pUTI89-like plasmid. Plasmid mapping was performed using a custom R script

available at github.com/maxlcummins/plasmidmapR. The presence of PAI-I_{CFT073} (NC_004431-P1), PAI-II_{CFT073} (NC_004431-P2), PAI-III_{536} (X16664), PAI-$IV_{APEC-O1}$ (NC_008563), and PAI-V_{536} (AJ617685) was assessed in a similar way to the plasmids, except the cut-offs were ≥ 95% of the length (scored as present), and < 95% but > 50% of the length (scored as partial).

4.6.5. Statistical analyses

To visualise the clustering of the isolates based on their virulence or resistance gene profile, gene presence/absence matrices were used in conjunction with classical (metric) multidimensional scaling (MDS) performed in R Studio using cmdscale and visualised using ggplot2 v3.3.0 (ggplot2.tidyverse.org/). To determine any statistically significant differences in gene presence-absence between UTI and non-UTI associated isolates Scoary 1.6.16 (46) was used with –no_pairwise flag, for not to infer any genes as causing group split but to identify any over/underrepresented genes.

Statistical tests were performed in R v4.1.2, X-square to determine the difference between plasmid types and *intI1* carriage performed using a standard chisq.test, while the Pairwise Wilcoxon test with Benjamini–Hochberg p-value correction for multiple testing was used to compare the number of ARG genes between different plasmid type groups utilising the pairwise_wilcox_test from the rstatix (CRAN.R-project.org/package=rstatix) package.

4.7. Results

4.7.1. Demography

This collection consisted of 336 *E. coli* draft genomes (mean contig size 46,363 bp, mean genome size 5,116,270 bp) originating from isolates sourced from the urine of patients attending OBH, Australia in 2006 (2006 isolates represent 95.5% of the total collection, n = 321). Most isolates were sourced from patients with urinary tract disease, including kidney infections (20.8%, n = 70) and lower UTIs (64%, n = 215). Isolates (n = 14) from chronic renal failure and kidney stones accounted for 4.2% of the collection while only six isolates (1.7%) were from patients with sepsis. Thirty-one (9.2%) isolates were sourced from patients with diseases not related to the urinary tract. Isolates sourced from females dominated the collection (n = 286) compared to males (n = 50), with 13 patients pregnant at the time of sampling. Patient age varied from less than one month of age to 97 years old, with an average of 53.9 years. Fifteen (4.5%) isolates were acquired from patients < 36-months-old and 141 (42%) from patients > 65 years. Relevant metadata is provided in Appendix 1.

4.7.2. Phylogeny

A maximum-likelihood core genome-based phylogenetic tree (Figure 4.1A) saw the isolates cluster according to Clermont phylogroup (47). EzClermont typing showed the presence of 10 clades with isolates within phylogroup B2 dominating the collection: A (n = 23, 6.8%), B1 (n = 23, 6.8%), B2 (n = 238, 70.8%), C (n = 4, 1.2%), D (n = 29, 8.6%), E (n = 2, 0.6%), F (n = 5, 1.4%), G (n = 3, 0.9%), U (n = 8, 2.4%), cryptic (n = 1, 0.3%). Ninety-one STs were identified. The pandemic STs dominated the collection led by ST73 (n = 54, 16.1%) followed by ST95 (n = 36, 10.1%), ST127 (n = 31, 9.2%), ST131 (n = 25, 7.4%), ST12 (n = 14, 4.2%), ST69 (n = 13, 3.9%), and ST144 (n = 12, 3.5%). Isolates with ST10 and ST550 (n = 10, 3% each), ST420 and ST80 (n = 7, 2.1% each), ST349, ST538 and ST58 (n = 5, 1.5% each) were also notable. The most common STs associated with specific uro-clinical syndromes are presented in Table 4.1. ST73 with serotype O6:H1 (n = 20) was the most prominent ST for patients diagnosed with lower UTIs. ST73 was also the most common ST in non-urinary tract related diseases, though the serotype differed (O6:H31 n=5). ST131 (serotype O25b:H4) followed by ST73 were frequently recovered from kidney infections. Four of the six isolates recovered from patients diagnosed with sepsis were ST95 and all carried a ColV F virulence plasmid.

Table 4.1: Most common STs and serotypes in isolates from different sources.

Uro-clinical syndrome	Most common STs	Most common serotypes
Kidney infections	ST131 (n = 9), ST73 (n = 8), ST95 (n = 7), ST127(n = 4) and ST420 (n = 4).	H4:O25b (n = 9), H31:O134-Gp6 (n=5), H31:O6 (n = 5), H1:O6 and H7:O1 (n = 4 each)
Lower UTIs	ST73 (n = 35), ST95 (n = 19), ST127 (n = 18), ST131 (n = 13), and ST144 (n = 8).	H1:O6 (n = 20), H31:O6 (n = 18), H7:O1 (n = 11), H4:O25b (n = 10) and H5:O75 (n = 9)
Non-urinary tract related (e.g. myocardial infarction, pneumonia, migraine)	ST73 (n = 6), ST127 (n = 5), ST10, ST12 and ST95 (n = 3 each).	H31:O6 (n = 5), H1:O2-O50-Gp7 and H18:O17-O77-Gp9 (n = 2 each)
Other urinary tract related (e.g. renal failure, nephrectomy, renal transplant)	ST127 (n = 3), ST131, 69 and 73 (n = 2 each).	H31:O6 (n = 3) and H4:O25b (n = 2)
Sepsis	ST95 (n = 4), ST131 and ST73 (n = 1 each)	H4:O25b (n = 2), H7:O1 (n = 2), H1:O2-O50-Gp7 and H4:O2-O50-Gp7 (n = 1 each)

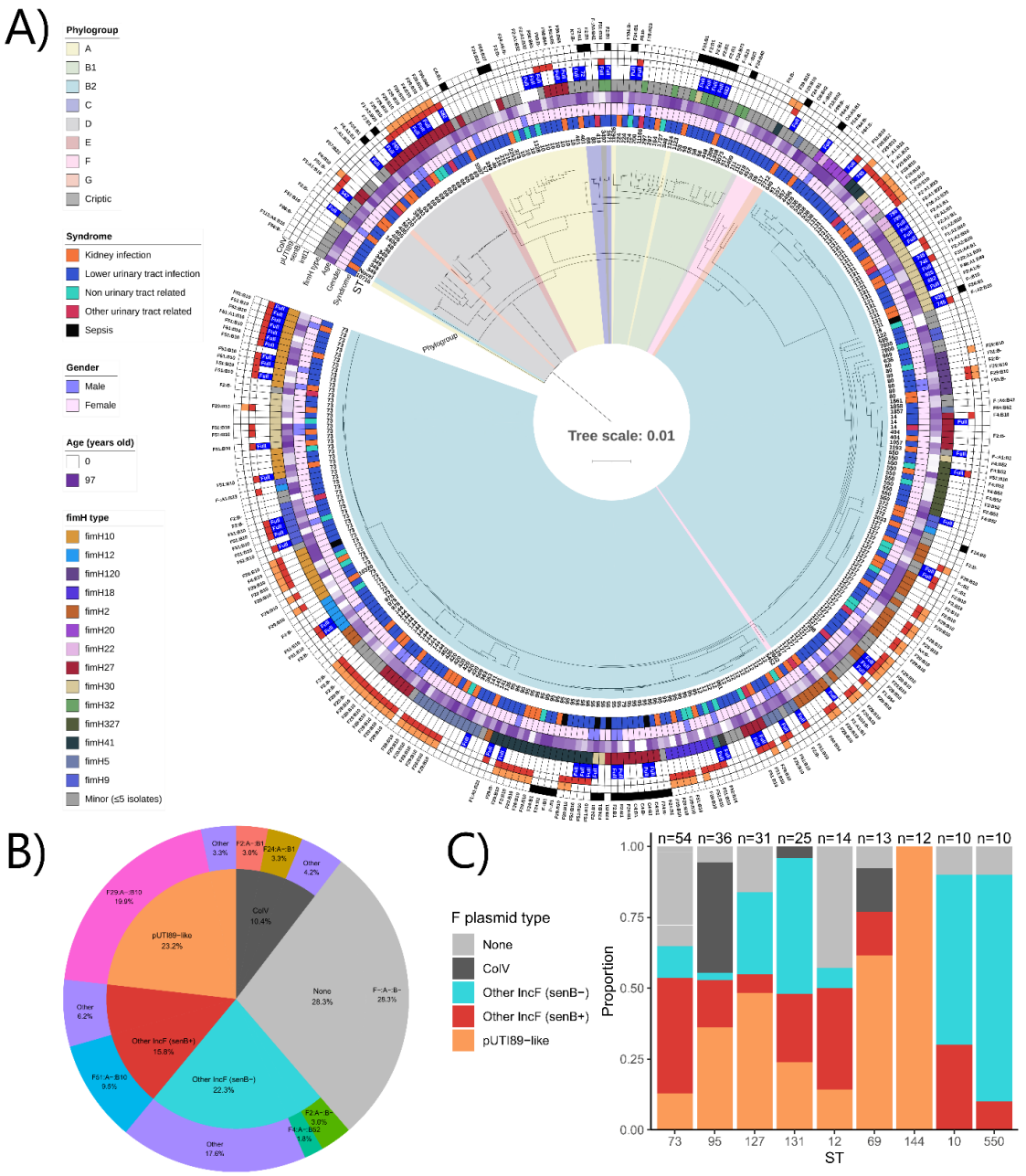


Figure 4.1: OBH isolate phylogeny and plasmid carriage. a) Maximum-likelihood phylogenetic tree of 336 ExPEC isolates sourced from OBH. Tree scale bars represent the number of substitutions per site of alignment. Core genome SNP tree rooted on outgroup *Escherichia fergusonii* ATCC 35469T (not shown), including 2631 genes with a total alignment length of 2,466,507 bp, constructed by IQ-TREE 2. Branch highlights coloured according to phylogroup identified by EzClermont, leaf names represent the ST of each isolate. The outer ring represents the presence of the ColV plasmid according to Liu criteria (45) (black). Second outer ring: indicates presence of pUTI89-like plasmid (orange). Third outer ring: presence of *senB* (red). Forth outer ring: presence of *intI1* (blue) and white text over the box represents truncation length of *intI1*. F RSTs are displayed at the periphery. **b)** Pie chart showing the distribution of F plasmid types across the whole collection, with dark grey for ColV positive isolates according to modified Liu et al., (2018) criteria (45), orange for isolates positive for pUTI89-like plasmids, red for isolates positive for senB but negative for pUTI89-like plasmids, and blue for isolates positive for F plasmids but negative for senB, pUTI89 or ColV, as well as main incF RSTs present in these

categories, with other for RSTs present in less than five isolates in particular categories. **c)** Bar chart of proportions of the same plasmid type categories for STs with ten or more isolates.

F virulence plasmids that carry *cjrABC* together with *senB* and those that carry ColV virulence markers are associated with distinct pandemic ExPEC lineages (18, 21, 48). F virulence plasmids influence *E. coli* host range, zoonotic potential (18, 21, 31, 49), and AMR carriage in ExPEC (18, 48). In this collection of 336 isolates from the OBH, 131 isolates (39%) carried F plasmids with *senB-cjrABC* and of these, 78 (59.5%) were F29:A-:B10 (n = 67; 86%) and 11 (14%; 5 RSTs) were closely related to pUTI89 according to the selection criteria. There were 53 (40.5%) *senB-cjrABC* carrying F plasmids with 16 RST types that are structurally different to pUTI89 with F51:A-:B10 dominating this group (n = 32; 60.4%). ColV F virulence plasmids with diverse RST types were detected in 35 (10.4%) genomes and were represented by 12 STs with ST95 the predominant lineage (n = 14; 40%). F24:A-:B1 was the most common (n = 11; 31.4%) ColV RST type. F plasmids were not detected in 95 (28.3%) of the 336 *E. coli* genomes and 75 (22.3%) *E. coli* genomes carry F plasmids without *senB-cjrABC* or were not ColV. Of these 75, 10 were ST131 clade C isolates of which there were only 15 in the collection. These data highlight the dominance of *E. coli* carrying F plasmids with *senB-cjrABC* virulence genes in urinary tract disease. Notably, *senB-cjrABC* positive plasmids were present in all 12 ST144 isolates and most of ST69 (77%, n = 10) and ST127 (55%, n = 17) isolates (Figure 4.1C).

E. coli with ST73, ST95, ST127, and ST131 accounted for nearly half (43%) of the collection. Additional core genome phylogenies were constructed for each of these four STs by including publicly available *E. coli* genomes with the same ST that originated from Australia. To provide a One Health perspective, genomes from non-human host origins were included and their relatedness to our clinical isolates was determined using SNP analyses.

4.7.2.1. ST73 phylogeny

ST73 is a serologically and phylogenetically diverse ST (50). ST73 isolates were the most abundant ST from the OBH collection. A recent antibiotic selection-free study conducted in the UK on a more contemporary collection further emphasises the continued dominance of ST73 within the ExPEC population (51). A separate core genome maximum-likelihood phylogenetic tree was constructed using ST73 isolates from this collection, all ST73 genomes available on EnteroBase from Australia (human n = 84, wild animal n = 1, and companion animal n = 30), and 24 ST73 genomes from companion animals as described in (52, 53) (Figure 4.2A). This phylogenetic analysis showed that companion animal-sourced isolates (canine and feline) tended to cluster together and that these isolates were less likely to carry F plasmid replicons (i.e., 28% of isolates originating from humans were F plasmid negative, as opposed to 68% in

isolates from animals, X-squared = 27.74, df = 1, p-value = 1.4e-7). Isolates from OBH were broadly distributed with no evident clustering by source or serotype (Figure 4.2A). Nevertheless, we did identify multiple instances of cases with small SNP differences in isolates originating from different hospitals as well as between clinical isolates and those derived from companion animals (Figure 4.2B). Specifically, we highlighted four clusters of ST73 isolates that displayed very close SNP distances: i) an isolate retrieved from a wild animal (flying fox, SRR11080153) which differed by 27 SNPs to companion animal-sourced isolates SRR14629706 and SRR14629726. The flying fox-sourced isolate also differed by 29 and 22 SNPs respectively to human-sourced isolates ERR2228592 and SRR11495787; ii) two closely related companion animal-sourced isolates SRR14629706 and SRR14629726 to human-sourced isolates SRR11495741 and SRR11495787 (average 20 SNPs, range 10 to 28); iii) A cluster of closely related human-sourced isolates, including six OBH isolates from different patients comprising four with lower UTI, one with renal calculi and one with acute myocadiac infarct, with an average of 24 SNPs across them, range 5-31 SNP; and iv) one feline companion animal isolate (Q13/1/48) and human isolate (ERR5355842) differed by 29 SNPs.

We were able to identify F plasmid replicon STs in 117 (61%) ST73 isolates depicted in Figure 4.2. ST73 was dominated by F plasmids carrying *senB-cjrABC* (n = 86, 73.5%) with only five isolates (4.3%) predicted to carry a ColV plasmid including a single ST73 carrying F24:A-:B1 and a cluster of four phylogenetically related isolates with an F untypable replicon ST. pUTI89 (F29:A-:B10 n = 28) ST73 isolates were dispersed across the phylogeny. F plasmids related to pUTI89 comprised F2:A-:B- (n = 2), F2:A-:B10 and F4:A-:B10 (both n = 1). *senB-cjrABC* plasmids distinct from pUTI89 with F51:A-:B10 (50/86; 58.1%) was the dominant RST type. A significant number (n = 76; 39%) of ST73 genomes did not carry an F plasmid (F-:A-:B-). Liu *et al*., (2018) developed criteria that have been useful in identifying *E. coli* carrying ColV plasmids. We noted that 66 ST73 isolates tested false positive using the Liu *et al*., (2018) ColV marker criteria many of which (n = 38, 58%) were F-:A-:B- by pMLST. Read mapping analyses failed to provide evidence of the presence of ColV plasmids in these isolates. Rather these false-positive ColV ST73 isolates carried a suite of genomic islands (see “Virulence-associated genes” section below) noted for carriage of ColV virulence markers. The false positives were due to the presence of ColV-associated genes on genomic islands rather than plasmids. being present in both plasmids and genomic islands. To address this, we increased the stringency of the Liu *et al*., (2018) criteria (44) to ≥ 95% identity and ≥ 95% gene coverage as outlined in the Methods and this removed the false positive status of these ST73 isolates.

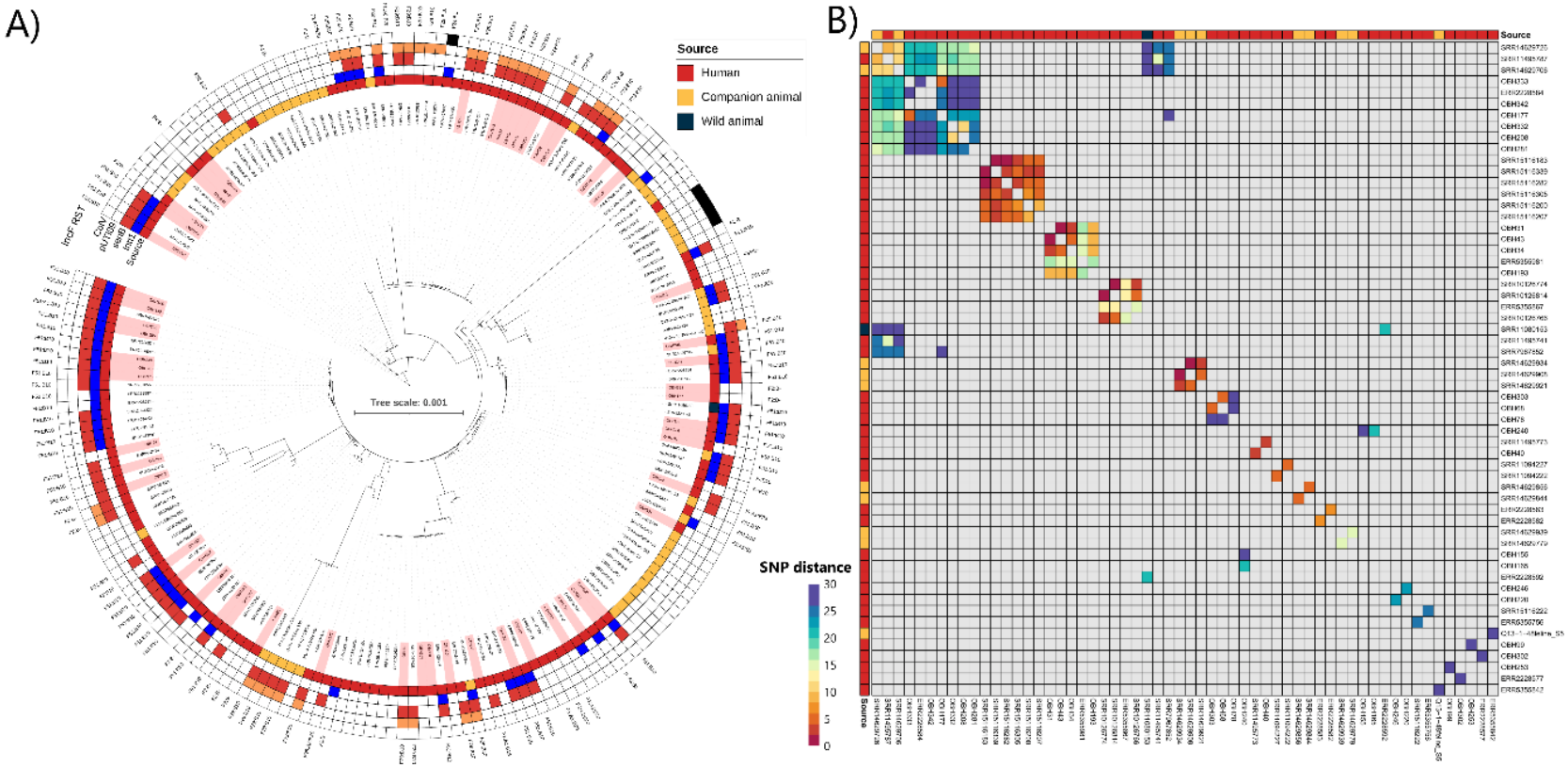


Figure 4.2: Phylogeny and SNP analysis of ST73 isolates. A) Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST73 isolates (alignment length 3,406,524 bp). Leaf names correspond to isolate names, highlighted by red colour isolate names are from this study. Inner ring represents the source of the isolates, second ring represents the presence of *intI1* (blue). Third ring indicates the presence of *senB* (red) with the fourth ring indicating the presence of the pUTI89-like plasmid (orange). The outer ring represents the presence of the ColV plasmid according to modified Liu criteria (45) (black). Outside of the rings, F RSTs are shown at the periphery. **B)** Pairwise SNP distance heatmap showing isolates with 30 SNPs (purple) or < 30 (purple to red colour scale). Values higher than the threshold or self-comparisons are greyed out. A pairwise SNP comparison heatmap, with a threshold of 100 SNPs, is available in Appendix 2.

4.7.2.2. ST95 phylogeny

A core genome maximum-likelihood phylogenetic tree of 36 ST95 isolates from the OBH as well as an additional 51 genomes from EnteroBase derived from Australian ST95 sourced from poultry (n = 25), humans (n = 22), and the environment (n = 4) with varying Hierarchical Clustering (HC20) designations was constructed. Previously we showed that *E. coli* ST95 isolates can be categorised into ten distinct clades (A-J) (18). A Maximum-Likelihood core genome phylogenetic tree (Figure 4.3A) revealed that the majority of ST95 isolates from the OBH collection belong to the human-centric clade A (n = 17, 47%), followed by clade B (n = 8, 22%), and clade E (n = 5, 14%). No OBH isolates were identified as belonging to clades F, H, or J. Small SNP differences between isolates from different sources were only observed between one human isolate and a poultry isolate (both in Clade I; 64 SNPs) (Figure 4.3B). However, examples of close human-sourced ST95 isolates from clade A were found including isolates from the same patients, OBH266 (cystitis) and OBH267 (cystitis) which were collected on the same day differed by 2 SNPs; OBH282 (nephritis) and OBH295 (sepsis) were 3 SNPs apart; OBH110 (cystitis) with OBH315 (cystitis) which were collected on the same day differed by 6 SNPs. Also present were closely related isolates from different patients: OBH284 (chronic cystitis) and OBH98 (urethritis) were 2 SNPs apart; and OBH12 (urethritis), OBH158 (septicemia), and SRR7967850 (sepsis) were an average of 15 SNPs apart (range 6 - 21). Notably, isolate SRR7967850 was sourced from Concord Repatriation Hospital in Sydney in 2013, approximately seven years and over 200 km apart.

pUTI89-like plasmids were found to be present in 13 isolates (36%), most of which belonged to clade A (n = 8, 61%) with *fimH*41 and serotype O1:H7. Similarly, ColV plasmids were present in 14 isolates (39%), however as expected most ColV positive isolates did not belong to human-dominated clades A or B (n = 10, 77%) but were dominant in clades I and also in clades C, E, and G. The distribution of ColV and pUTI89-like plasmids across the major STs identified in this study can be viewed in Figure 4.1C. Notably, ST95 had the highest carriage of ColV plasmids, while other STs including ST127, ST69, and ST144 were predominantly associated with pUTI89-like plasmid carriage. Consistent with this we recently described the carriage of pUTI89-like plasmids in ST127 from canine and human origins (31).

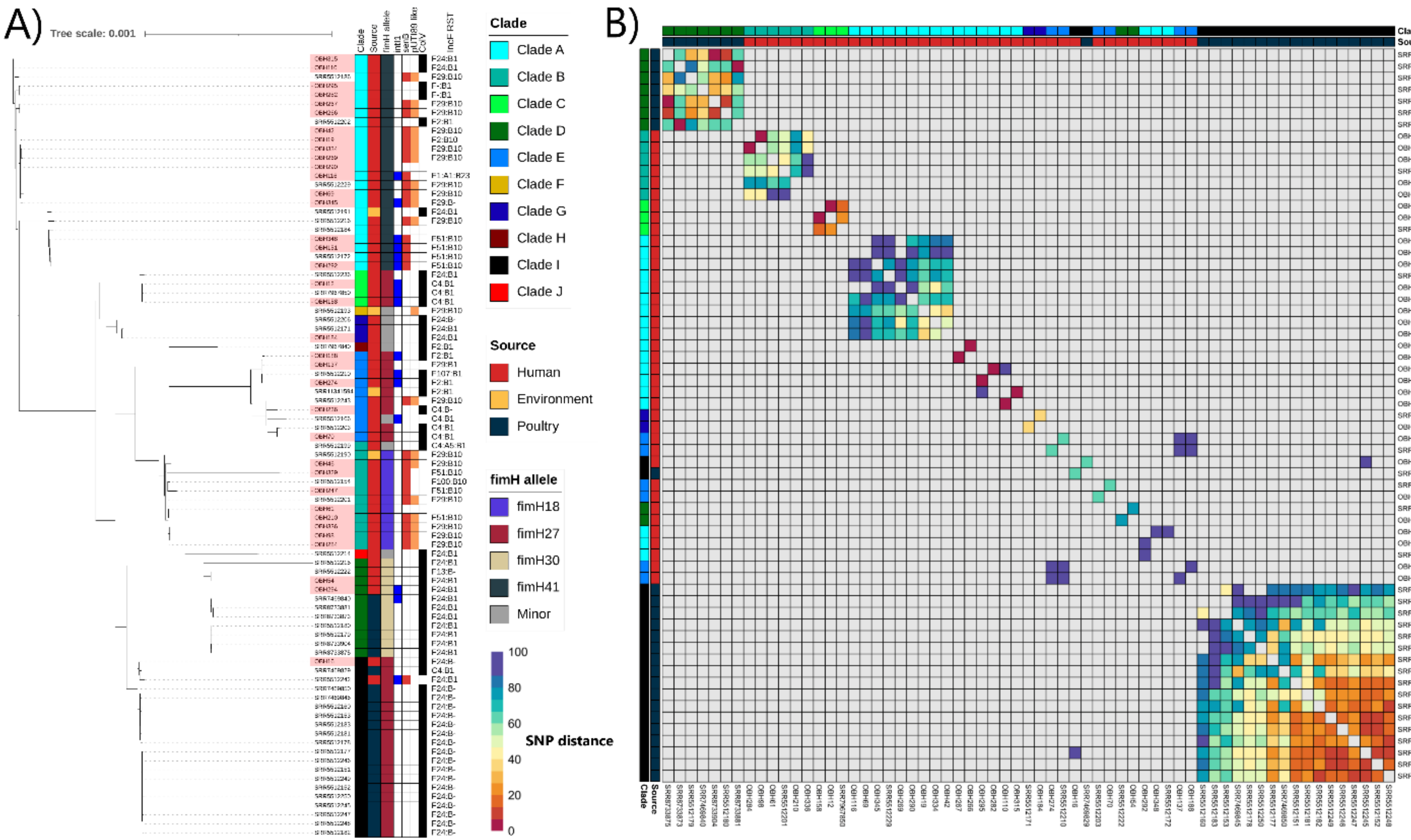

A)
Tree scale: 0.001
B)
Clade
Clade A
Clade B
Clade C
Clade D
Clade E
Clade F
Clade G
Clade H
Clade I
Clade J
Source
Human
Environment
Poultry
fimH allele
fimH18
fimH27
fimH30
fimH41
Minor
100
80
60
40
20
0
SNP distance

Figure 4.3: Phylogeny and SNP analysis of ST95 isolates. a) Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST95 isolates from this collection, plus 51 additional Australian ST95 isolates of varying HC20. Total alignment length: 3,582,702 bp. Leaf names correspond to isolate names, isolates from this collection are highlighted in red. Coloured squares on the right represent metadata for each isolate, first column represents ST95 clades, second refers to source, third column represents *fimH* allele, fourth: presence of *intI1* (blue), fifth: presence of *senB* (red), sixth: presence of pUTI89 like plasmid (orange). Outer ring represents presence of the ColV plasmid according to Liu criteria (45) (black). Outside label corresponds to F RST of the isolate. **b)** Pairwise SNP distance heatmap showing pairs with 100 SNPs (purple) or lower (red), values higher than threshold or self-comparisons are greyed out.

4.7.2.3. ST127 phylogeny

A core genome maximum-likelihood phylogenetic tree of ST127 was built using an additional 41 (human n = 18, companion animal n = 22, wild animal n = 1) Australian ST127 isolates from EnteroBase (Figure 4.4). The 31 ST127 isolates from this collection were broadly distributed among four main clusters. Overall, ST127 isolates were found to be diverse, with 6,178 SNPs separating the two most distant isolates (companion animal-sourced isolates SRR14629755 and SRR14629640). However, we did identify several closely related isolates (Figure 4.4B) including: i) a human-sourced (OBH136) isolate with two companion animal isolates SRR14629715 and SRR14629890 (94 and 97 SNPs); ii) three companion animal isolates in *fimH*283-H2 cluster (SRR14629715, SRR14629890 and SRR14629874) have an average 87 SNPs (range 69 - 103) with four clinical isolates (OBH230,164,57,235) from OBH; iii) five closely related isolates (average 34 SNPs, range: 18 - 60) all originating from one patient of OBH (OBH187, 239, 262, 272, SRR11495782) that were collected over a two month period with the uro-clinical syndrome changing from pyelonephritis to cystitis to glomerulonephritis; iv) three isolates from three pathological sites: pyrexia of unknown origin (OBH85), cystitis (OBH343) and urosepsis (SRR11495794), and from different patients were found to have average distance of 40 SNPs (range 18-53); and v) two isolates, OBH189 and OBH224, isolated 24 days apart from the same patient with a recurrent UTI were indistinguishable (0 SNPs).

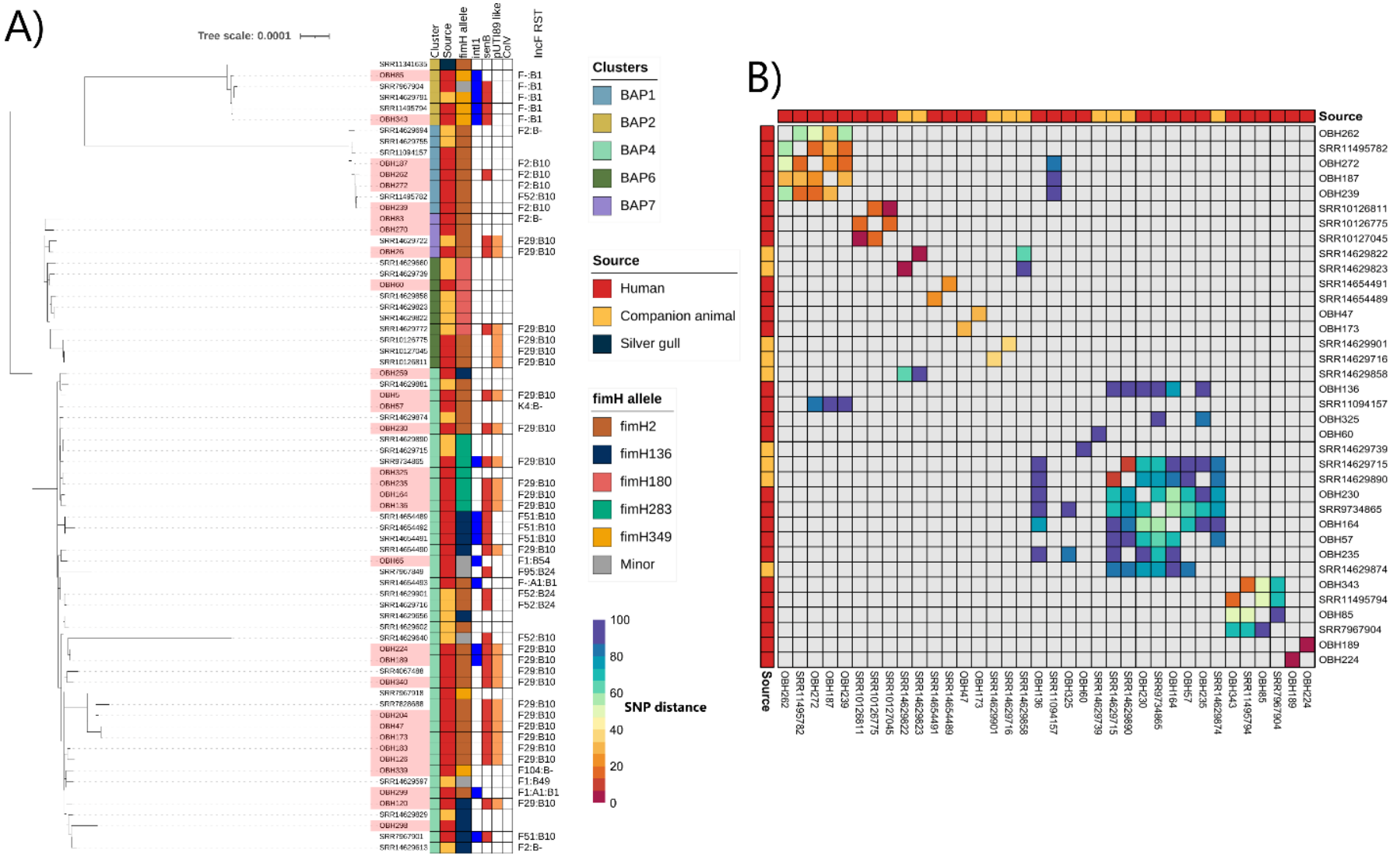

A)
Tree scale: 0.0001
Cluster
Source
fimH allele
intI1
senB
pUTI89 like
ColV
IncF RST
Clusters
BAP1
BAP2
BAP4
BAP6
BAP7
Source
Human
Companion animal
Silver gull
fimH allele
fimH2
fimH136
fimH180
fimH283
fimH349
Minor
100
80
60
40
20
0
SNP distance
B)
Source
OBH262
SRR11495782
OBH272
OBH187
OBH239
SRR10126811
SRR10126775
SRR10127045
SRR14629822
SRR14629823
SRR14654491
SRR14654489
OBH47
OBH173
SRR14629901
SRR14629716
SRR14629858
OBH136
SRR11094157
OBH325
OBH60
SRR14629739
SRR14629715
SRR14629890
OBH230
SRR9734865
OBH164
OBH57
OBH235
SRR14629874
OBH343
SRR11495794
OBH85
SRR7967904
OBH189
OBH224

Figure 4.4: Phylogeny and SNP analysis of ST127 isolates. a) Mid-point rooted core genome-based maximum-likelihood phylogenetic tree of ST127 isolates. Rooted on ST372 isolate (SRA accession number: SRX10969655), not shown on the tree. Total alignment length of 3,579,832 bp. Leaf names correspond to isolate names, with isolates from this study highlighted in red. Coloured squares on the right represent metadata for isolate: first column represents source, second column *fimH*, third: presence of *intI1* (blue), fourth: presence of pUTI89 like plasmid (orange) and fifth: presence of *senB* (red). Outer ring represents presence of the ColV according to Liu criteria (46) (black). Outside label corresponds to incF RST of the isolate. **b)** Pairwise SNP distance heatmap showing pairs with 100 SNPs (purple) or <100 SNPs (purple to red colour gradient), values higher than threshold or self-comparisons are greyed out.

4.7.2.4. ST131 phylogeny

We recently performed comprehensive genomic analyses on all available *E. coli* ST131 genomes originating from Australia (n = 284) (54) (Chapter 6), so here we provide a core genome-based maximum-likelihood tree of ST131 isolates of this collection plus eight Australian ST131 isolates representing the major clades C1, C2, B and A (54, 55) (Figure 4.5A). Most isolates from this collection belonged to clade C (n = 15, 60%), followed by seven (28%) in clade B and three (12%) in clade A. The diversity of F RSTs differed among ST131 clades: OBH isolates in clades A and B carried three different F RSTs F29:B10 (n = 6), F2:A1:B23 (n = 2), and F-:A1:B23 (n = 2) and all of these were pUTI89-like plasmids. Clade C1 isolates also had three types present (F2:A1:B1 n = 3, F1:A2:B20 n = 3, F36:A1:B20 n = 1) while in clade C2 every isolate (n = 8) had a unique F plasmid RST.

ST131 isolates are known to be highly clonal, especially within the C1 and C2 clades (54), and here we found examples of low pairwise SNP counts across all clades (Figure 4.5B). Three isolates from the same patient differed by average SNP count of 33 (range 28-36 SNPs). These isolates were collected one and two months apart, the first two with a kidney infection (OBH179 and OBH216) and the third isolate from a patient with a lower UTI (OBH297). Different patient-sourced isolates in clade A (OBH214 and OBH226) sampled five days apart showed a SNP count of 23 while clade B (OBH96 and OBH124) isolates collected 17 days apart differed by 55 SNPs. Only one ST131 isolate from our collection was found to be ColV positive (OBH346, clade C2).

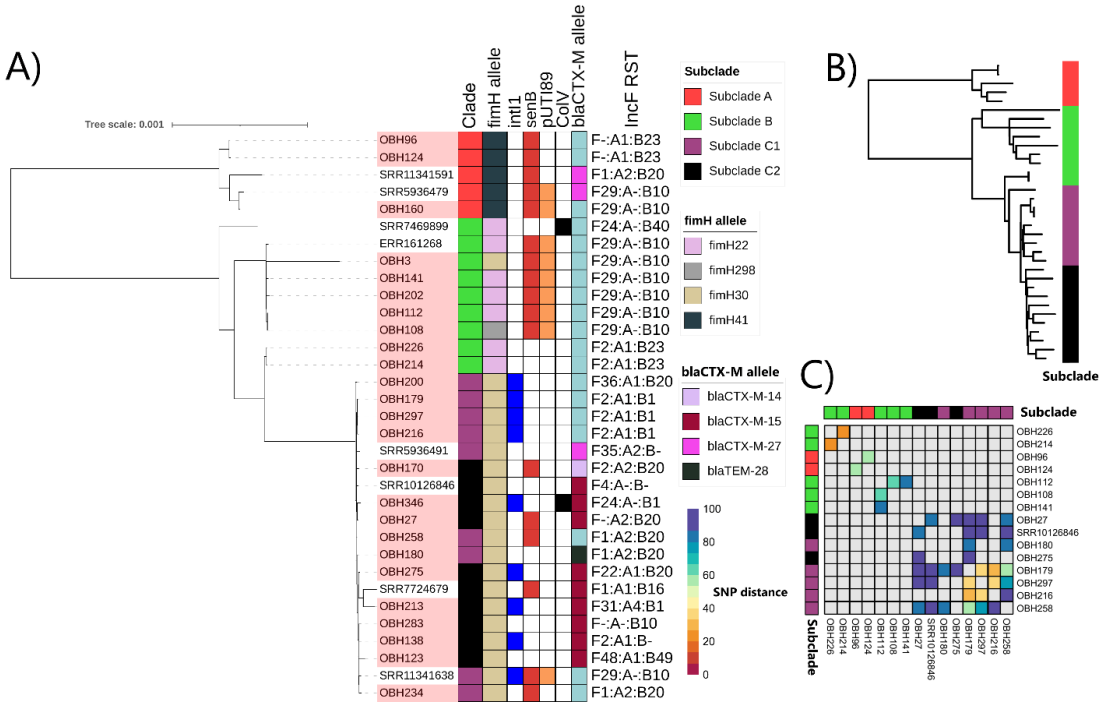


Figure 4.5: Phylogeny and SNP analysis of ST131 isolates. a) Mid-point rooted core genome-based Maximum-Likelihood phylogenetic tree of ST131 isolates. Total alignment length of 3,697,845 bp. Leaf names correspond to isolate names, isolates from the current study are highlighted in red. Coloured squares on the right represent metadata for isolate, first column: clade, second column: *fimH* alleles, third: *intI1* (blue), fourth: presence of pUTI89 like plasmid (orange), fifth: presence of *senB* (red). Outer column represents the presence of the ColV plasmid according to Liu criteria (45) (black). Outside label corresponds to incF RST. **b)** Reference (EC958) based mid-point rooted recombination filtered core SNP Maximum-Likelihood phylogenetic tree of constructed to resolve clades C1 and C2. **c)** Pairwise SNP distance heatmap showing pairs with 100 SNPs (purple) or < 100 SNPs (purple to red colour gradient), with values higher than threshold or self-comparisons greyed out.

4.7.3. Antimicrobial resistance

A comprehensive analysis of the collection revealed the presence of 95 distinct ARGs within this collection (Figure 4.6A; Appendix 3). Only 25 isolates (7.4%) carried genes encoding ESBLs, and these were predominantly found in ST131 isolates (n = 9). Only isolates in ST131 clade C carried ESBLs with seven positive for $bla_{CTX\text{-}M\text{-}15}$, and one for $bla_{CTX\text{-}M\text{-}14}$ and $bla_{CTX\text{-}M\text{-}27}$ each. Sulfonamide resistance genes were found in more than a third of all isolates (36%; n = 122), and the most common gene was *sul1* (n = 73), followed by *sul2* (n = 63) and *sul3* (n = 4). Similarly, genes conferring resistance to aminoglycosides were identified in 31% of isolates (n = 105), the three most common being *aph(3'')-Ib* (*strA*) (n = 54), *aph(6)-Id* (*strB*) (n = 38), and *ant(3'')-IIa* (n = 29). Genes conferring resistance to trimethoprim were found in approximately a fifth of this collection (19%; n = 64), the most prevalent being *dfrA17* (n = 24), *dfrA5* (n = 14), and *dfrA1* (n = 10). The recently characterised *dfrB4* gene (56) was identified in two isolates (ST448 OBH273 and ST1193 OBH152). Tetracycline resistance genes were identified in 47 (14%) isolates, with

tet(A) (n = 26) being most common, followed by *tet(B)* (n = 21) and *tet(D)* (n = 1). Only one isolate carried both *tet(A)* and *tet(B)*. Genes conferring resistance to macrolides, chloramphenicol, streptothricin, quinolones, and fosfomycin were less common. The macrolide resistance gene *mphA* was found in 7.7% (n = 26) of isolates. Genes conferring chloramphenicol resistance were found in seven isolates (2.1%) (*cmlA1*: n = 5, *cmlA5*: n = 1, *catB3*: n = 1 each). Streptothricin resistance gene *sat2* was found in seven isolates (two ST73, one of ST12, ST14, ST59, ST144, and ST372 each). The quinolone resistance gene *qnrD1* was identified in only one ST127 isolate. Similarly, just one ST1727 isolate carried the fosfomycin resistance gene *fosA7*. The collection was also screened for specific point mutations in *gyrA* and *parC* genes that confer resistance to fluoroquinolones. The *gyrA*-1AB mutation was identified in 5.6% of the isolates (n = 19), most commonly in ST131 (n = 14), but also in ST393 (n = 2), ST448, ST405 and ST1193 (n = 1 each). The *gyrA*-1AB and *parC*-1aAB dual mutations were identified in 4.5% of isolates (n = 15) – 14 in ST131 (Clades C1 and C2) isolates and in one ST405 isolate.

Overall, the carriage of ARGs was distributed as follows: 3.9% of isolates for gentamycin resistance (based on *aac(3)-IId*, *aac(3)-IIe*, *aph(3')-IIa* (57)), 36% for sulfonamide resistance alone, 18% for trimethoprim and sulfonamide resistance combined, 7% for ESBLs, and 4% for fluoroquinolone resistance. Fifty-nine percent (n = 199) of isolates did not carry genes or mutations conferring resistance to any of these antibiotics classes and no isolates carried carbapenemase genes.

Despite the skewed frequency of antibiotic resistance genes overall, there were no associations between uro-clinical syndrome and ARG profiles and nor any statistically significant differences between uro-clinical syndrome and the mean number of ARGs. Similarly, there were no associations between ARG profiles or the mean number of ARGs in terms of patients biological sex and age (Appendix 4 and 5). The distribution of ARGs conferring resistance to clinically important antibiotics differed among the major STs in this collection, with ST131, ST69, and ST10 isolates exhibiting robust MDR profiles (Figure 4.6B).

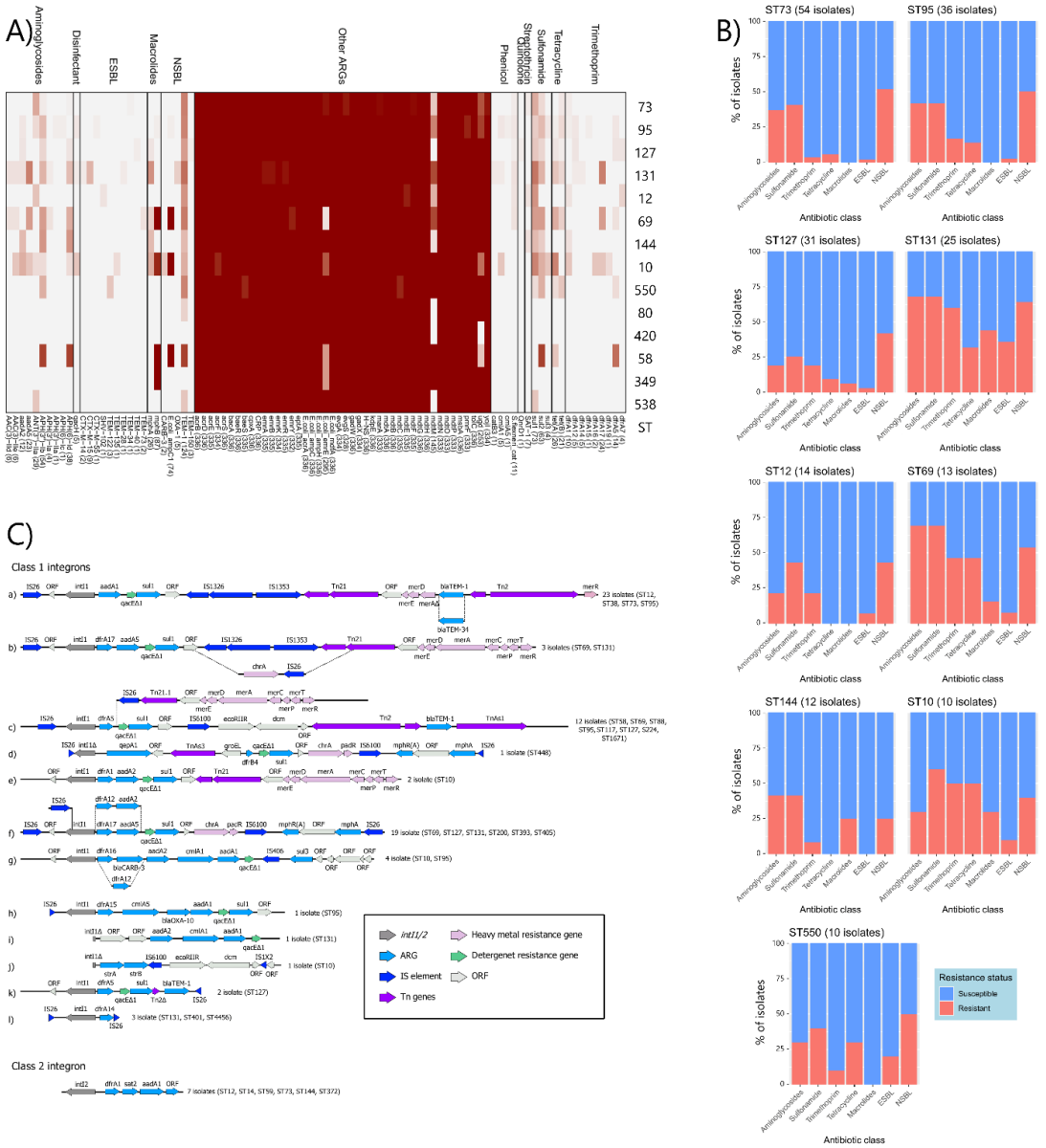


Figure 4.6: **Genomic antibiotic resistance analysis. a)** Gene screening heat map for ARGs for STs with ≥ 5 isolates. **b)** Bar chart of genotypic AMR in STs with ≥ 5 isolates. **c)** Class 1 (a–l) and class 2 integrons structures identified in this collection. Integron integrase genes in grey, ARG in cyan, IS elements in blue, Tn genes in purple, heavy metal resistance genes in pink, detergent resistance genes in green, other ORFs in white.

Class 1 integrons are low-cost structures that can capture and express ARGs and lead to an MDR phenotypes. The capture and shuffling of ARGs are mediated via the integrase *intI1* (58). Here, a total of 89 isolates (26.5%) contained *intI1*. Most (n = 74) carried the non-truncated form of *intI1*. In the remaining 15 isolates, eight different *intI1* truncations were observed (Figure 4.1), the most prevalent being $intI1_{\Delta 745}$ (n = 5). Overall, the most common class 1 integron structure was *intI1-aadA1-qacEΔ1-sul1* associated with a Tn*21* transposon, and a Tn*2* transposon (containing the bla_{TEM-1} gene) inserted into the *mer* operon (mercury resistance) (Figure 4.6Ca). This particular class 1 integron confers resistance to aminoglycosides, sulphonamides, NSBLs,

and quaternary ammonium compounds. This structure was dominant in ST73 (n = 16/23) and isolates from patients with lower UTIs (n = 17/23). In one isolate (ST73), a SNP in the bla_{TEM-1} gene gave rise to bla_{TEM-34}, thereby providing putative ESBL resistance. Only one isolate (an ST131 sourced from a kidney infection) was identified carrying two class 1 integrons (Figure 4.6Ci, 4.6Cl). The highest prevalence of class 1 integrons was found in ST131 (60%, n = 15) followed by ST10 (50%, n = 5) and ST69 (46%, n = 6), which correlates with these STs having MDR profiles as mentioned above. There was a significant difference found between plasmid type and class 1 integron carriage (X-squared = 83.46, df = 4, p-value < 2.2e-16), with *senB-cjrABC*+/non-pUTI89-like plasmids (encompassing 17 IncF RSTs; most common F51:A-:B10) being most associated with *intI1*. In particular, 84% (n = 27) of isolates carrying F51:A-B10 were *intI1* positive (Appendix 3). Conversely, isolates that carried pUTI89-like plasmids had significantly lower *intI1* carriage (n = 6, 7.7% vs n = 70, 27%, X-squared = 7.99, df = 1, p-value = 0.0047). Isolates that carried *senB-cjrABC*/non-pUTI89-like plasmids also carried on average the most ARGs associated with HGT (n = 5), which was significantly different to other plasmid types (pUTI89-like p-value = 5.3e-08; *senB*- IncF plasmids p-value = 0.000422; no IncF plasmid p-value = 2.32e-17) (Appendix 5).

The Class 2 integron *intI2-dfrA1-sat2-aadA1* was identified in seven isolates (Figure 4.6C), providing resistance to trimethoprim, streptothricin, and aminoglycosides. Other integron classes were not detected.

4.7.4. Virulence-associated genes (VAGs)

A total of 218 VAGs were identified in the collection (Figure 4.7A; Appendix 3). The molecular definition of UPEC is defined as the presence of ≥ 3 of the following genes: *chuA*, *fyuA*, *yfcV*, and *vat* (59). By this definition, 230 (60%) isolates qualified as UPEC, however, 124 isolates obtained from patients with UTIs did not possess ≥ 3 of these genes, including 82 isolates from lower UTIs and 29 from kidney infections. Notably, a Scoary analysis found no statistically significant difference in VAG carriage (or any other gene) between UTI- and non-UTI-associated isolates (Appendix 6). While no associations between the overall VAG profile and uro-clinical syndrome, sex and age were observed, we did find that isolates originating from children (0-12 yrs) had a significantly higher mean VAG count compared to isolates derived from adults (18-64 yrs) and elders (64+ yrs) (Appendix 4 and 5). Across all age groups, sexes, and pathologies, genes involved in iron acquisition, adhesion, immune evasion, and toxins were widespread (61). The ferric yersiniabactin uptake receptor *fyuA* was found in 86% (n = 288) of isolates. The iron-regulatory proteins *irp1* (n = 285 85%), and *irp2* (n = 283, 84%) were similarly common, with the complete yersiniabactin siderophore operon *ybtAEPQSTYX* present in 85% (n = 286) of isolates. The

Salmochelin siderophore system encoded by *iroBCDEN* was identified in 46% (n = 153) isolates. The full aerobactin operon *iucABCD* and *iutA* were present in 12% (n = 39) of isolates, while 41% of isolates (n = 137) carried the *iucABCD* operon but lacked *iutA*.

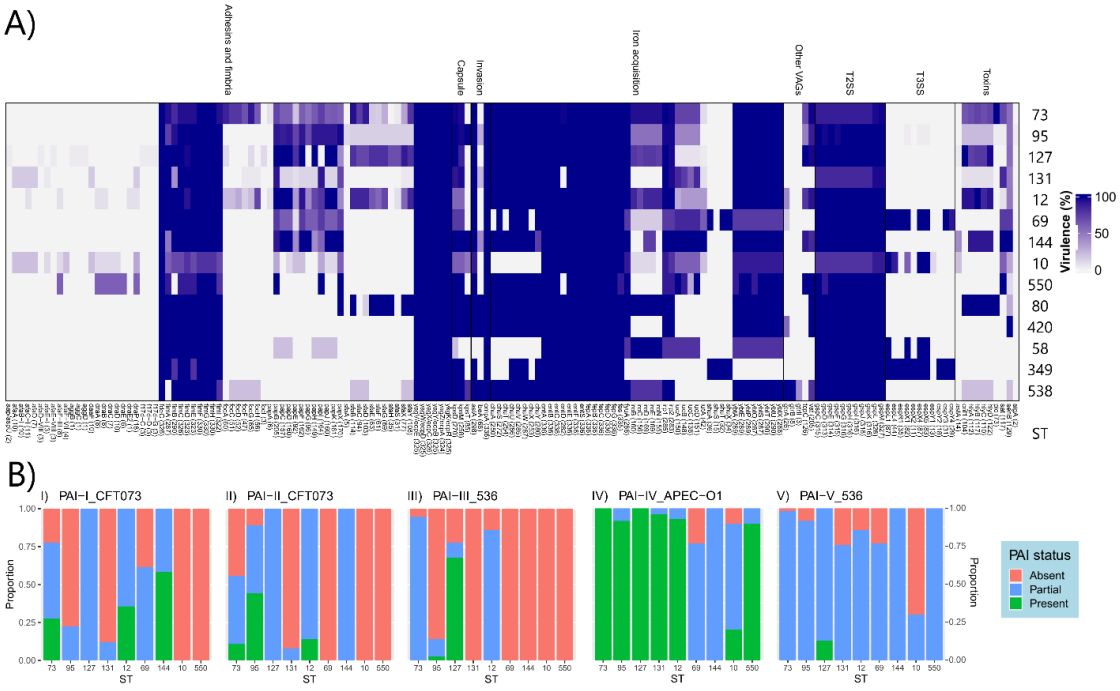


Figure 4.7: Genomic virulence analysis. a) Gene screening heat map for VAGs for STs with ≥ 5 isolates. Abbreviations used: T2SS – Type 2 Secretion System, T3SS – Type 3 Secretion System. **b)** Presence status of five PAI for STs with ≥ 10 isolates.

Specificity for uroepithelium has been demonstrated for type 1 fimbriae and P-fimbriae which are important for UTI-causing ExPEC (60). Type I fimbriae genes were present in most isolates and *fimH*, which encodes a protein that binds to uroepithelial associated α-D-mannosylated proteins, was present in 98% of isolates (n = 330). P-fimbriae, encoded by the *pap* operon, bind to receptors located in the upper urinary tract and are associated with nephritis (61). However, we did not find a higher association between the presence of pap operon genes and kidney infections.

ExPEC have well documented immune evasion mechanisms and produce toxins facilitating damage of the host tissues (62, 63). Immune evasion mechanisms include increased serum survival protein (*iss*) present here in 75% (n = 252) of isolates, VirB5-like protein TraT (*traT*) observed in 65.2% of isolates (n = 219), and outer-membrane protein T (*ompT*) detected in 10.4% of isolates (n = 35). Toxins identified included enterotoxin TieB protein (*senB*) observed in 37.5% (n = 126) of isolates, full alpha-hemolysin operon (*hlyABCD*) was present in 31.8% (n =

107) of isolates, cytotoxic necrotizing factor 1 (*cnf1*) was found in 30.9% of isolates (n = 104) and heat-stable enterotoxin 1 (*astA*), was identified in 4.2% (n = 14) of isolates.

Iron acquisition systems, adhesins, and toxins are commonly found on PAIs. We screened for the presence of five PAIs (Appendix 7) – PAI-I_{CFT073} carrying the hemolysin operon (*hlyABCD*), PAI-II_{CFT073} with P fimbriae operon (*papACDEFGHIJK*), PAI-III_{536} containing adhesion (*sfaABCDEFGHSX*) and the iron acquisition genes (*iroBCDEN*, putative hemin receptor), PAI-$IV_{APEC-O1}$ which carries the yersiniabactin operon (*fyuA, irp1*, *irp2*, *ybtAEPQSTUX*), and PAI-V_{536} containing adhesins (*pixABCDFGHJ*, *pgtABCP*) and capsule related operons (*kpsCEFMSTU*).

Evidence for a complete PAI-I_{CFT073} was found in 29 isolates, mostly in ST73 isolates (n = 15) and ST144 (n = 7). A partial copy of PAI-I_{CFT073} (< 95% identity and > 50% coverage) was found in 107 isolates. Most isolates (65%; n = 70) carrying a partial copy of PAI-I_{CFT073} contained the full *hly* operon but not a full *pap* operon, lacking *papG* in 31 (32%) genomes and *papA* in all genomes. Twenty-nine isolates also did not possess the *shiA*-like gene. PAI-II_{CFT073} was identified in 26 isolates, mostly with ST95 (n = 16), followed by ST73 (n = 6). Nearly half (44%; n = 24) of ST73, 44% (n = 16) of ST95, 85% (n = 12) of ST12 and all ST127 and ST144 isolates carried a partial island, with most missing *ireA* (63%, n = 74) and *papA* (76%, n = 89). PAI-III_{536} was present in 23 isolates, almost exclusively within ST127 isolates (n = 21), while ST73 contained the most (94%; n = 51) partial versions of this PAI followed by ST12 (86%; n = 12). Isolates carrying a partial PAI-III_{536} were missing the putative lysine/cadaverine transporter and the putative cadaverine decarboxylase genes. Cadaverine has been shown to promote UPEC colonisation of the bladder (64). PAI-$IV_{APEC-O1}$ was present in most isolates (n = 205; 61%). The first 28,321 bp (36.5%) of this PAI has 97.92% identity to the complete HPI. Most phylogroup B2 isolates carried either full PAI-$IV_{APEC-O1}$ (n = 189) or a partial version (n = 46). Of the 46 partial versions, 40 encompassed a complete HPI. Only four isolates contained full PAI-V_{536}, and all of them are ST127. Only a small proportion (13%, n = 32) of phylogroup B2 isolates are negative for PAI-V_{536} compared to other phylogroups (54%, n = 52).

In summary, the top six pandemic STs (ST73, ST131, ST95, ST69, ST127, and ST12) often carried the five PAIs to different degrees. PAI-$IV_{APEC-O1}$ was present in almost all B2 phylogroup isolates. Out of all the STs, ST127 isolates had the highest proportion of carrying either complete or partial versions of all five PAIs.

4.8. Discussion

This study provided an in-depth characterisation of a large collection (n = 336) of urine-sourced clinical *E. coli* isolates from patients attending an Australian regional hospital in 2006. Through

genomic analyses, several key observations were made. First, *E. coli* ST73, ST95, ST127, and ST131 dominated the collection. Second, an overall low ARG carriage was detected but was higher in ST131, ST69, and ST10 isolates, and isolates that carried *senB*+/non-pUTI89-like IncF plasmids. Third, several clinical isolates were closely related to both companion and wild animal-sourced isolates. Fourth, there were episodes of closely related isolates recovered from the same patient over an extended period and associated with different pathologies. Fifth, there were multiple episodes of closely related isolates from different patients in the hospital and with patients from other hospitals in Australia and sixth, no associations between specific VAGs and pathologies, though there was a distinct distribution of PAIs within dominant pandemic STs. An important limitation of this study was the lack of temporal association between the human clinical and animal isolates, given that the OBH collection is from 2006 and the animal isolates used for comparison were all isolated later (between 2009 – 2019).

ExPEC are comprised of diverse STs; however, globally only a subset is responsible for most infections. A recent systematic review found that 85% of ExPEC infections were attributed to just 20 STs, the top five being ST131 > ST69 > ST10 > ST405 > ST38 (6); however, here the most common STs were ST73 > ST95 > ST127 > ST131 > ST12 which accounted for 47% of all isolates. Most of the 20 STs flagged by Manges et al. (2019) (65) were represented in this collection, except for ST648, ST354, ST167, and ST23. While ExPEC STs vary in abundance and diversity within human populations (66), our observed difference between the top ST distributions likely stemmed from the fact that most studies included in the review were biased because the collections were based on antibiotic resistance (6). Indeed, our previous study on trimethoprim-resistant *E. coli* UTI isolates from the same hospital (14) (Chapter 5) found that the most common STs were ST131 and ST69. However, when sample selection was not based on AMR phenotypes, our current results reflect other studies from the United Kingdom, the United States of America, and Canada, wherein ST73, ST95, ST127, and ST131 were the most frequently isolated STs from large cohorts of UTIs (67-69).

Given the importance of AMR, ongoing surveillance of resistant populations undoubtedly has merit, but it does not provide an accurate portrayal of the STs most responsible for UTIs. *E. coli* ST73 as a primary causative agent of UTIs is likely to be underrepresented in the literature as the lineage was often described as pan-sensitive to antibiotics (70, 71) and plasmid naïve (72). Not so in this collection, as we found 52% of ST73 isolates carried ARGs conferring resistance to β-lactams, 41% to sulphonamides, 37% to aminoglycosides, 35% carried class 1 integron integrase *intI1*, and 77% carried at least one plasmid replicon. Notably however only 2% (n=1) of ST73 isolates displayed resistance to ESBLs. The most common plasmid type detected in ST73

isolates was *senB-cjrABC*+/non-pUTI89-like IncF plasmids with RST (F51:B10). We found this plasmid type was significantly associated with higher ARG carriage, while pUTI89-like plasmids (also *senB-cjrABC*+), which were dominated by F29:A-:B10 in our collection, were associated with lower ARG carriage. This is somewhat unsurprising given that this plasmid lineage has been reported to be associated with pan-susceptible ExPEC strains (49).

Despite the positive association of *senB-cjrABC*+/non-pUTI89-like IncF plasmids and ARGs, the overall ARG carriage in this retrospective collection was mostly low, particularly regarding ARGs conferring resistance to antibiotics commonly prescribed for complicated UTIs such as gentamycin (4% of isolates), ESBLs (7%), and fluoroquinolones (4%). However, ARGs encoding resistance to trimethoprim/sulphonamide (19%) were considerable. The majority of ESBL genes (mainly $bla_{\text{CTX-M-15}}$) were found in ST131, which is unsurprising given that this lineage is thought to have played a central role in the global increase of ESBL-producing *Enterobacteriales* (73). Likewise, most of the genotypic fluoroquinolone resistance we observed was attributed to dual *parC* and *gyrA* mutations in ST131 isolates. Interestingly, however, while dual *gyrA*-1AB and *parC*-1aAB were previously thought to be highly specific to ST131 clade C isolates (74), we found these specific mutations in an ST405 isolate (OBH263, from 72 y.o. female with kidney infection). Recently we also reported these mutations in ST131 clade A isolates (55). Overall, ST131 isolates possessed the most robust MDR profiles, followed by ST10 and ST69 isolates. These STs had the highest prevalence of *intI1*, giving credence to class 1 integrons as reliable markers for MDR (75). The most common class 1 integron structure in this collection was *intI1-aadA1-qacEΔ1-sul1*, which has been identified in *E. coli* previously, as well as in numerous other bacterial species including *Aeromonas*, *Bacillus*, *Citrobacter*, *Klebsiella*, *Pseudomonas*, *Salmonella* and *Vibrio* (76). Curiously, while truncated *intI1* genes are common among *E. coli* populations (77) and can be used as epidemiological markers (78), here only 2% of *intI1* genes had deletions, which may speak to the retrospective nature of this collection, or possibly that truncated *intI1* genes are more prevalent in non-human isolates. For example, in an Australian study of 425 critically drug-resistant *E. coli* from gulls sampled in 2012, 242 (57%) were determined to carry class 1 integrons. Of these, 64% showed 3´-truncations in *intI1*, most often associated with IS26, with identical truncations found across multiple lineages. Often these truncations were missed by high-throughput gene identification.

Recent studies have shown that rural populations are more likely to receive inappropriate antibiotics for inappropriate durations than urban populations (79, 80). OBH services regional, rural, and remote communities, so one could speculate that the carriage of ARGs and class 1 integrons in UTI isolates would be higher here than in an urban hospital setting. However, the

overall *intI1* carriage was lower compared to that reported in a study of UTI isolates from three metropolitan Australian hospitals during the same time period (26% vs 34%) (81). One possible explanation for this is that urban surface waters and sediments can have higher ARG loads and plasmid carriage compared to rural samples (82), meaning that the urban environment could be a significant driver for AMR.

UTIs are predominately community-acquired and, beyond human gastrointestinal tracts, the environment, sewage and abattoir waste, retail meats, and animals (wild, food, and companion) have all been suggested as ExPEC reservoirs (18, 21, 27, 28, 32, 46, 50, 78). To explore potential interspecies movement, we performed phylogenetic and SNP analyses on our top four STs (ST73, ST95, ST131, ST127) encompassing genomes of human and non-human-sourced isolates of Australian origin. ST73 was previously described as a human-specific ExPEC lineage (83, 84). However, recent studies have shown that ST73 is prominent in cats (54, 85, 86), dogs (50, 53, 87), killer whales (88), and some avian populations (89). Consistent with previous studies, our phylogenetic analysis showed ST73 as diverse (51, 90), and that animal-sourced isolates tended to cluster together (53, 54). Nevertheless, we did identify four clusters of very closely related ST73 isolates, including two human-sourced isolates that were 22 and 29 SNPs from a flying fox-sourced isolate, and a human-sourced isolate and cat-sourced isolate separated by 29 SNPs. These low inter-host SNP counts are particularly relevant given the phylogenetic diversity of ST73 and suggest that this clonal group is well-adapted to a range of host species. We previously showed the potential for interspecies movement in an Australian ST131 population (55), however, ST131s have a relatively conserved core genome even amongst the different clades. Conversely, ST73 isolates can differ by 1000s of SNPs (90) and indeed we found that some human-sourced ST73 isolates differed from other ST73 human-sourced isolates by up to 5766 SNPs. We also found low SNP distances between ST127 isolates from human and animal origins. *E. coli* ST127 has been described as an emerging, highly virulent, human pathogen (67, 91) but has also been isolated from companion animals (32, 50, 53, 54, 85), killer whales (88), and bats (92). We found six instances of our clinical isolates being separated by < 100 SNPs from five companion animal-sourced isolates. While our analysis only included animal isolates from Australia, a recent ST127 study showed geographically proximal and distal human- vs companion animal-sourced isolates differed by < 30 SNPs (32). *E. coli* ST95 is a prominent cause of both human and avian diseases (93), and the zoonotic potential of this lineage is well documented, particularly in relation to O1:H7 strains (94-96), or those belonging to clade I (18). We found that though the most common serotype in this hospital ST95 population was O1:H7, these were dispersed predominately among human-dominated clade A. Only one of our clinical isolates

belonged to the Australian poultry-dominated clade I and was 64 SNPs from the closest poultry isolate and both isolates carried ColV plasmids.

It is evident that certain STs are more likely to cause UTIs, however, the mechanisms behind pathogenesis remain elusive. Congruent with previous studies we found that ST73 was most commonly isolated from patients with cystitis (83), ST95 was associated with sepsis (97), and ST131 was most commonly retrieved from kidney infections (98). But we found no significant difference in VAG carriage between isolates from different pathologies. Indeed, we found no significant differences in VAGs, or any other gene, between UTI isolates and non-UTI isolates. Furthermore, we found several examples of isolates with small SNPs distances in isolates from patients experiencing different pathologies. For example, six ST73 isolates from different patients (4 cystitis, 1 renal calculus, and 1 acute myocadiac infarct) differed by an average of 24 SNPs (range 5-31 SNPs). Even the molecular definition of UPEC is problematic. By definition, 60% of our isolates qualified as UPEC but excluded were 124 isolates from patients with UTIs, including 82 isolates from lower UTIs and 29 from kidney infections. The majority of ExPEC VAGs encompass siderophore systems, urinary tract specific adhesins and immune evasion effectors (99), all of which were ubiquitous in our collection. It is important to note that *E. coli* belonging to phylogroup B2, particularly ST73, ST131, ST95, and ST141 are dominant in the gut of healthy humans and their frequency has increased in the commensal faecal population over the past 40 years (100, 101). These genes prime ExPEC for survival outside of the gastrointestinal tract but also play a role in its persistence in the human gut, while pathogenesis is likely to involve underlying patient factors, such as comorbidities and age (102), or a combination of VAGs. On that note, PAIs can encode several different VAGs and are known to play a major role in the evolution of ExPEC (103). We observed a correlation between five specific PAIs and STs and found that overall ST127 isolates carried the most PAIs, a feature that has been noted in a recent report of ST127 (32).

Our study had some limitation, in addition to the previously mentioned lack of temporal data, the presence of PAIs was predicted using short read mapping, which identifies gene presence, but these genes are not necessarily within the same genomic context as the screened PAIs. Future long-read sequencing studies are needed to provide greater insight into these initial observations. Lastly, AMR phenotypes were not available for the isolates, however ARG presence and resistance phenotypes are typically highly congruent in *E. coli* (104).

4.9. Conclusions

This study describes a phylogenomic analysis of a large cohort of *E. coli* from the urine of patients with uro-centric disease from a major rural hospital in NSW, Australia. While it is unsurprising that the top six STs are the pandemic lineages ST73 > ST95 > ST127 > ST131 > ST12 which accounted for 47% of all isolates, our study overlays the distribution of F virulence plasmids, pathogenicity islands, and antibiotic resistance gene cargo and provides deeper insights into lineage evolution. The collection is dominated by *E. coli* strains that carry plasmids with *senB-cjrABC*. Among these strains, those carrying F29:A-:B10 (pUTI89) exhibit lower carriage of antibiotic resistance. However, *E. coli* strains carrying *senB-cjrABC* genes but not closely related to pUTI89 tended to have a greater antibiotic resistance gene load and were typically *intI1*+. We were able to provide evidence of isolate movements between patients within and across hospital settings, demonstrate the persistence of a clonal lineage from the same patient over short and considerable time periods, and importantly demonstrate evidence of occasional interspecies transmission, particularly between humans and companion animals. Our study underscores the importance of taking a One Health genomic approach to pathogen surveillance.

4.10. References

1. Abbo LM, Hooton TM. Antimicrobial Stewardship and Urinary Tract Infections. Antibiotics (Basel). 2014;3(2):174-92. doi: 10.3390/antibiotics3020174.
2. AURA 2019: Third Australian report on antimicrobial use and resistance in human health.: Australian Commission on Safety and Quality in Health Care (ACSQHC) Sydney: ACSQHC; 2019 [Available from: https://www.safetyandquality.gov.au/sites/default/files/2019-06/AURA-2019-Report.pdf.
3. McCall B. Artificial intelligence bridges the complexity of intractable medical problems. The Lancet Digital Health. 2021;3(1):e8-e9. doi: 10.1016/S2589-7500(20)30296-X.
4. Poolman JT, Wacker M. Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. The Journal of Infectious Diseases. 2016;213(1):6-13. doi: 10.1093/infdis/jiv429.
5. Manges AR. Escherichia coli causing bloodstream and other extraintestinal infections: tracking the next pandemic. The Lancet Infectious Diseases. 2019;19(12):1269-70. doi: 10.1016/S1473-3099(19)30538-9.
6. Manges Amee R, Geum Hyun M, Guo A, Edens Thaddeus J, Fibke Chad D, Pitout Johann DD. Global Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. Clinical Microbiology Reviews. 2019;32(3):e00135-18. doi: 10.1128/CMR.00135-18.
7. Zilberberg MD, Shorr AF. Secular trends in gram-negative resistance among urinary tract infection hospitalizations in the United States, 2000-2009. Infect Control Hosp Epidemiol. 2013;34(9):940-6. doi: 10.1086/671740.
8. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. J Antimicrob Chemother. 2017;72(8):2145-55. doi: 10.1093/jac/dkx146.

9. Wesolek JL, Wu JY, Smalley CM, Wang L, Campbell MJ. Risk Factors for Trimethoprim and Sulfamethoxazole-Resistant Escherichia Coli in ED Patients with Urinary Tract Infections. Am J Emerg Med. 2022;56:178-82. doi: 10.1016/j.ajem.2022.03.052.
10. Carmona-Cartaya Y, Hidalgo-Benito M, Borges-Mateus LM, Pereda-Novales N, González-Molina MK, Quiñones-Pérez D. Community-Acquired Uropathogenic Escherichia coli, Antimicrobial Susceptibility, and Extended-Spectrum Beta-Lactamase Detection. MEDICC Rev. 2022;24(2):20-5. doi: 10.37757/mr2022.v24.n2.2.
11. García-Meniño I, Lumbreras P, Lestón L, Álvarez-Álvarez M, García V, Hammerl JA, et al. Occurrence and Genomic Characterization of Clone ST1193 Clonotype 14-64 in Uncomplicated Urinary Tract Infections Caused by Escherichia coli in Spain. Microbiol Spectr. 2022:e0004122. doi: 10.1128/spectrum.00041-22.
12. Bean DC, Livermore DM, Papa I, Hall LM. Resistance among Escherichia coli to sulphonamides and other antimicrobials now little used in man. J Antimicrob Chemother. 2005;56(5):962-4. doi: 10.1093/jac/dki332.
13. Tano ZN, Kobayashi RK, Candido EP, Dias JB, Perugini LF, Vespero EC, et al. Susceptibility to first choice antimicrobial treatment for urinary tract infections to Escherichia coli isolates from women urine samples in community South Brazil. Braz J Infect Dis. 2022:102366. doi: 10.1016/j.bjid.2022.102366.
14. Li D, Reid CJ, Kudinha T, Jarocki VM, Djordjevic SP. Genomic analysis of trimethoprim-resistant extraintestinal pathogenic Escherichia coli and recurrent urinary tract infections. Microbial Genomics. 2020. doi: 10.1099/mgen.0.000475.
15. Manges Amee R, Johnson James R, Mulvey Matthew A, Stapleton Ann E, Klumpp David J. Reservoirs of Extraintestinal Pathogenic Escherichia coli. Microbiology Spectrum. 2015;3(5):3.5.06. doi: 10.1128/microbiolspec.UTI-0006-2012.
16. Nesporova K, Wyrsch Ethan R, Valcek A, Bitar I, Chaw K, Harris P, et al. Escherichia coli Sequence Type 457 Is an Emerging Extended-Spectrum-β-Lactam-Resistant Lineage with Reservoirs in Wildlife and Food-Producing Animals. Antimicrobial Agents and Chemotherapy. 2020;65(1):e01118-20. doi: 10.1128/AAC.01118-20.
17. Tarabai H, Wyrsch ER, Bitar I, Dolejska M, Djordjevic SP. Epidemic HI2 Plasmids Mobilising the Carbapenemase Gene blaIMP-4 in Australian Clinical Samples Identified in Multiple Sublineages of Escherichia coli ST216 Colonising Silver Gulls. Microorganisms. 2021;9(3). doi: 10.3390/microorganisms9030567.
18. Cummins ML, Reid CJ, Djordjevic SP. F Plasmid Lineages in Escherichia coli ST95: Implications for Host Range, Antibiotic Resistance, and Zoonoses. mSystems. 2022;7(1):e0121221. doi: 10.1128/msystems.01212-21.
19. Kantele A, Lääveri T, Mero S, Häkkinen IMK, 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. 2019;70(2):210-8. doi: 10.1093/cid/ciz182.
20. Stephens C, Arismendi T, Wright M, Hartman A, Gonzalez A, Gill M, et al. F Plasmids Are the Major Carriers of Antibiotic Resistance Genes in Human-Associated Commensal Escherichia coli. mSphere. 2020;5(4):e00709-20. doi: 10.1128/mSphere.00709-20.
21. Reid CJ, Cummins ML, Börjesson S, Brouwer MSM, Hasman H, Hammerum AM, et al. A role for ColV plasmids in the evolution of pathogenic Escherichia coli ST58. Nature Communications. 2022;13(1):683. doi: 10.1038/s41467-022-28342-4.

22. Johnson Timothy J, Elnekave E, Miller Elizabeth A, Munoz-Aguayo J, Flores Figueroa C, Johnston B, et al. Phylogenomic Analysis of Extraintestinal Pathogenic Escherichia coli Sequence Type 1193, an Emerging Multidrug-Resistant Clonal Group. Antimicrobial Agents and Chemotherapy. 2019;63(1):e01913-18. doi: 10.1128/AAC.01913-18.
23. Dobrindt U. (Patho-)Genomics of Escherichia coli. International Journal of Medical Microbiology. 2005;295(6):357-71. doi: 10.1016/j.ijmm.2005.07.009.
24. Galardini M, Clermont O, Baron A, Busby B, Dion S, Schubert S, et al. Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus Escherichia revealed by a genome-wide association study. PLOS Genetics. 2020;16(10):e1009065. doi: 10.1371/journal.pgen.1009065.
25. Cusumano Corinne K, Hung Chia S, Chen Swaine L, Hultgren Scott J. Virulence Plasmid Harbored by Uropathogenic Escherichia coli Functions in Acute Stages of Pathogenesis. Infection and Immunity. 2010;78(4):1457-67. doi: 10.1128/IAI.01260-09.
26. Reid CJ, McKinnon J, Djordjevic SP. Clonal ST131-H22 Escherichia coli strains from a healthy pig and a human urinary tract infection carry highly similar resistance and virulence plasmids. Microbial Genomics. 2019;5(9). doi: 10.1099/mgen.0.000295.
27. 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). doi: 10.1099/mgen.0.000250.
28. Moran RA, Holt KE, Hall RM. pCERC3 from a commensal ST95 Escherichia coli: A ColV virulence-multiresistance plasmid carrying a sul3-associated class 1 integron. Plasmid. 2016;84-85:11-9. doi: 10.1016/j.plasmid.2016.02.002.
29. Nolan LK, Barnes HJ, Vaillancourt J-P, Abdul-Aziz T, Logue CM. Colibacillosis. Diseases of Poultry2013. p. 751-805.
30. Hicks SJ, Rowbury RJ. Virulence plasmid-associated adhesion of Escherichia coli and its significance for chlorine resistance. Journal of Applied Bacteriology. 1986;61(3):209-18. doi: 10.1111/j.1365-2672.1986.tb04278.x.
31. Elankumaran P, Browning GF, Marenda MS, Reid CJ, Djordjevic SP. Close genetic linkage between human and companion animal extraintestinal pathogenic Escherichia coli ST127. Current Research in Microbial Sciences. 2022;3:100106. doi: 10.1016/j.crmicr.2022.100106.
32. Gaio D, Anantanawat K, To J, Liu M, Monahan L, Darling AE. Hackflex: low-cost, high-throughput, Illumina Nextera Flex library construction. Microbial Genomics. 2022;8(1). doi: 10.1099/mgen.0.000744.
33. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90. doi: 10.1093/bioinformatics/bty560.
34. Lanfear R, von Haeseler A, Woodhams MD, Schrempf D, Chernomor O, Schmidt HA, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution. 2020;37(5):1530-4. doi: 10.1093/molbev/msaa015.
35. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3. doi: 10.1093/bioinformatics/btv421.
36. Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ. 2014;2:e243. doi: 10.7717/peerj.243.

37. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. Easy phylotyping of Escherichia coli via the EzClermont web app and command-line tool. Access Microbiol. 2020;2(9):acmi000143. doi: 10.1099/acmi.0.000143.
38. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9. doi: 10.1093/bioinformatics/btu153.
39. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46. doi: 10.1186/gb-2014-15-3-r46.
40. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother. 2017;72(10):2764-8. doi: 10.1093/jac/dkx217.
41. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2022;50(D1):D912-d7. doi: 10.1093/nar/gkab1107.
42. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48(D1):D517-d25. doi: 10.1093/nar/gkz935.
43. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903. doi: 10.1128/aac.02412-14.
44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi: 10.1186/1471-2105-10-421.
45. 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. doi: 10.1128/mBio.00470-18.
46. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biology. 2016;17(1):238. doi: 10.1186/s13059-016-1108-8.
47. Gonzalez-Alba JM, Baquero F, Cantón R, Galán JC. Stratified reconstruction of ancestral Escherichia coli diversification. BMC Genomics. 2019;20(1):936. doi: 10.1186/s12864-019-6346-1.
48. Stephens Craig M, Adams-Sapper S, Sekhon M, Johnson James R, Riley Lee W, Limbago Brandi M. Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. mSphere. 2017;2(2):e00390-16. doi: 10.1128/mSphere.00390-16.
49. Elankumaran P, Cummins ML, Browning GF, Marenda MS, Reid CJ, Djordjevic SP, et al. Genomic and Temporal Trends in Canine ExPEC Reflect Those of Human ExPEC. Microbiology Spectrum. 2022;10(3):e01291-22. doi: 10.1128/spectrum.01291-22.
50. 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). doi: 10.1099/mgen.0.000255.
51. Kallonen T, Brodrick HJ, Harris SR, et al. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131 [published online ahead of print, 2017 Jul 18]. Genome Res. 2017;27(8):1437-1449. doi:10.1101/gr.216606.116

52. Kidsley AK, O'Dea M, Saputra S, Jordan D, Johnson JR, Gordon DM, et al. Genomic analysis of phylogenetic group B2 extraintestinal pathogenic E. coli causing infections in dogs in Australia. Vet Microbiol. 2020;248:108783. doi: 10.1016/j.vetmic.2020.108783.
53. Kidsley AK, O'Dea M, Ebrahimie E, Mohammadi-Dehcheshmeh M, Saputra S, Jordan D, et al. Genomic analysis of fluoroquinolone-susceptible phylogenetic group B2 extraintestinal pathogenic Escherichia coli causing infections in cats. Vet Microbiol. 2020;245:108685. doi: 10.1016/j.vetmic.2020.108685.
54. Li D, Wyrsch ER, Elankumaran P, Dolejska M, Marenda MS, Browning GF, et al. Genomic comparisons of Escherichia coli ST131 from Australia. Microb Genom. 2021;7(12). doi: 10.1099/mgen.0.000721.
55. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, et al. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A. 2014;111(15):5694-9. doi: 10.1073/pnas.1322678111.
56. Toulouse JL, Edens TJ, Alejaldre L, Manges AR, Pelletier JN. Integron-Associated DfrB4, a Previously Uncharacterized Member of the Trimethoprim-Resistant Dihydrofolate Reductase B Family, Is a Clinically Identified Emergent Source of Antibiotic Resistance. Antimicrob Agents Chemother. 2017;61(5). doi: 10.1128/aac.02665-16.
57. Shaw KJ, Rather PN, Hare RS, Miller GH. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. Microbiol Rev. 1993;57(1):138-63. doi: 10.1128/mr.57.1.138-163.1993.
58. Lacotte Y, Ploy MC, Raherison S. Class 1 integrons are low-cost structures in Escherichia coli. Isme j. 2017;11(7):1535-44. doi: 10.1038/ismej.2017.38.
59. Spurbeck RR, Dinh PC, Jr., Walk ST, Stapleton AE, Hooton TM, Nolan LK, et al. Escherichia coli isolates that carry vat, fyuA, chuA, and yfcV efficiently colonize the urinary tract. Infect Immun. 2012;80(12):4115-22. doi: 10.1128/iai.00752-12.
60. Dale AP, Woodford N. Extra-intestinal pathogenic Escherichia coli (ExPEC): Disease, carriage and clones. J Infect. 2015;71(6):615-26. doi: 10.1016/j.jinf.2015.09.009.
61. Wright KJ, Hultgren SJ. Sticky fibers and uropathogenesis: bacterial adhesins in the urinary tract. Future Microbiol. 2006;1(1):75-87. doi: 10.2217/17460913.1.1.75.
62. Ostolaza H, Goñi FM. Interaction of the bacterial protein toxin alpha-haemolysin with model membranes: protein binding does not always lead to lytic activity. FEBS Lett. 1995;371(3):303-6. doi: 10.1016/0014-5793(95)00927-2.
63. Wiles TJ, Kulesus RR, Mulvey MA. Origins and virulence mechanisms of uropathogenic Escherichia coli. Exp Mol Pathol. 2008;85(1):11-9. doi: 10.1016/j.yexmp.2008.03.007.
64. Bower, J. M., Gordon-Raagas, H. B., & Mulvey, M. A. Conditioning of uropathogenic Escherichia coli for enhanced colonization of host. Infection and immunity. 2009;77(5), 2104–2112. doi: 10.1128/IAI.01200-08
65. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. Global Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. Clinical Microbiology Reviews. 2019;32(3):e00135-18. doi: doi:10.1128/CMR.00135-18.
66. Rodríguez I, Figueiredo AS, Sousa M, Aracil-Gisbert S, Fernández-de-Bobadilla MD, Lanza VF, et al. A 21-Year Survey of Escherichia coli from Bloodstream Infections (BSI) in a Tertiary Hospital Reveals How Community-Hospital Dynamics of B2 Phylogroup Clones Influence Local BSI Rates. mSphere. 2021;6(6):e00868-21. doi: doi:10.1128/msphere.00868-21.
67. Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia

coli from Northwest England. Journal of Antimicrobial Chemotherapy. 2010;67(2):346-56. doi: 10.1093/jac/dkr451.
68. Yamaji R, Rubin J, Thys E, Friedman CR, Riley LW, Diekema DJ. Persistent Pandemic Lineages of Uropathogenic Escherichia coli in a College Community from 1999 to 2017. Journal of Clinical Microbiology. 2018;56(4):e01834-17. doi: doi:10.1128/JCM.01834-17.
69. Fibke CD, Croxen MA, Geum HM, Glass M, Wong E, Avery BP, et al. Genomic Epidemiology of Major Extraintestinal Pathogenic Escherichia coli Lineages Causing Urinary Tract Infections in Young Women Across Canada. Open Forum Infectious Diseases. 2019;6(11). doi: 10.1093/ofid/ofz431.
70. Doumith M, Day M, Ciesielczuk H, Hope R, Underwood A, Reynolds R, et al. Rapid Identification of Major Escherichia coli Sequence Types Causing Urinary Tract and Bloodstream Infections. Journal of Clinical Microbiology. 2015;53(1):160-6. doi: doi:10.1128/JCM.02562-14.
71. Dale AP, Woodford N. Extra-intestinal pathogenic Escherichia coli (ExPEC): Disease, carriage and clones. Journal of Infection. 2015;71(6):615-26. doi: 10.1016/j.jinf.2015.09.009.
72. Bengtsson S, Naseer U, Sundsfjord A, Kahlmeter G, Sundqvist M. Sequence types and plasmid carriage of uropathogenic Escherichia coli devoid of phenotypically detectable resistance. Journal of Antimicrobial Chemotherapy. 2011;67(1):69-73. doi: 10.1093/jac/dkr421.
73. Pitout JD, DeVinney R. Escherichia coli ST131: a multidrug-resistant clone primed for global domination. F1000Res. 2017;6. doi: 10.12688/f1000research.10609.1.
74. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, et al. Abrupt Emergence of a Single Dominant Multidrug-Resistant Strain of Escherichia coli. The Journal of Infectious Diseases. 2013;207(6):919-28. doi: 10.1093/infdis/jis933.
75. Gillings MR. Integrons: Past, Present, and Future. Microbiology and Molecular Biology Reviews. 2014;78(2):257-77. doi: doi:10.1128/MMBR.00056-13.
76. Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics. 2009;25(8):1096-8. doi: 10.1093/bioinformatics/btp105.
77. Nemergut DR, Robeson MS, Kysela RF, Martin AP, Schmidt SK, Knight R. Insights and inferences about integron evolution from genomic data. BMC Genomics. 2008;9(1):261. doi: 10.1186/1471-2164-9-261.
78. Wyrsch ER, Nesporova K, Tarabai H, Jamborova I, Bitar I, Literak I, et al. Urban Wildlife Crisis: Australian Silver Gull Is a Bystander Host to Widespread Clinical Antibiotic Resistance. mSystems. 2022;0(0):e00158-22. doi: doi:10.1128/msystems.00158-22.
79. Clark AW, Durkin MJ, Olsen MA, Keller M, Ma Y, O'Neil CA, et al. Rural-urban differences in antibiotic prescribing for uncomplicated urinary tract infection. Infect Control Hosp Epidemiol. 2021;42(12):1437-44. doi: 10.1017/ice.2021.21.
80. Yau JW, Thor SM, Tsai D, Speare T, Rissel C. Antimicrobial stewardship in rural and remote primary health care: a narrative review. Antimicrobial Resistance & Infection Control. 2021;10(1):105. doi: 10.1186/s13756-021-00964-1.
81. Gündoğdu A, Long YB, Vollmerhausen TL, Katouli M. Antimicrobial resistance and distribution of sul genes and integron-associated intI genes among uropathogenic Escherichia coli in Queensland, Australia. Journal of Medical Microbiology. 2011;60(11):1633-42. doi: 10.1099/jmm.0.034140-0.

82. McInnes RS, uz-Zaman MH, Alam IT, Ho SFS, Moran RA, Clemens JD, et al. Metagenome-Wide Analysis of Rural and Urban Surface Waters and Sediments in Bangladesh Identifies Human Waste as a Driver of Antibiotic Resistance. mSystems. 2021;6(4):e00137-21. doi: doi:10.1128/mSystems.00137-21.
83. de Souza da-Silva AP, de Sousa VS, Martins N, da Silva Dias RC, Bonelli RR, Riley LW, et al. Escherichia coli sequence type 73 as a cause of community acquired urinary tract infection in men and women in Rio de Janeiro, Brazil. Diagnostic Microbiology and Infectious Disease. 2017;88(1):69-74. doi: 10.1016/j.diagmicrobio.2017.01.024.
84. 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 Pathogens and Disease. 2015;12(4):302-10. doi: 10.1089/fpd.2014.1860.
85. Bourne JA, Chong WL, Gordon DM. Genetic structure, antimicrobial resistance and frequency of human associated Escherichia coli sequence types among faecal isolates from healthy dogs and cats living in Canberra, Australia. PLOS ONE. 2019;14(3):e0212867. doi: 10.1371/journal.pone.0212867.
86. Liu X, Thungrat K, Boothe DM. Multilocus Sequence Typing and Virulence Profiles in Uropathogenic Escherichia coli Isolated from Cats in the United States. PLOS ONE. 2015;10(11):e0143335. doi: 10.1371/journal.pone.0143335.
87. Valat C, Drapeau A, Beurlet S, Bachy V, Boulouis H-J, Pin R, et al. Pathogenic Escherichia coli in Dogs Reveals the Predominance of ST372 and the Human-Associated ST73 Extra-Intestinal Lineages. Frontiers in Microbiology. 2020;11. doi: 10.3389/fmicb.2020.00580.
88. Melendez D, Roberts MC, Greninger AL, Weissman S, No D, Rabinowitz P, et al. Whole-genome analysis of extraintestinal pathogenic Escherichia coli (ExPEC) MDR ST73 and ST127 isolated from endangered southern resident killer whales (Orcinus orca). Journal of Antimicrobial Chemotherapy. 2019;74(8):2176-80. doi: 10.1093/jac/dkz159.
89. Becker SSA, van Vliet AHM, Stegger M, Johannesen TB, Semmler T, Cunha M, et al. Genomic analysis of the zoonotic ST73 lineage containing avian and human extraintestinal pathogenic Escherichia coli (ExPEC). Vet Microbiol. 2022;267:109372. doi: 10.1016/j.vetmic.2022.109372.
90. Alhashash F, Wang X, Paszkiewicz K, Diggle M, Zong Z, McNally A. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR Escherichia coli ST73: high levels of genomic and plasmid diversity in causative isolates. Journal of Antimicrobial Chemotherapy. 2015;71(2):339-43. doi: 10.1093/jac/dkv365.
91. Darling AE, McKinnon J, Worden P, Santos J, Charles IG, Chowdhury PR, et al. A draft genome of Escherichia coli sequence type 127 strain 2009-46. Gut Pathogens. 2014;6(1):32. doi: 10.1186/1757-4749-6-32.
92. Nowak K, Fahr J, Weber N, Lübke-Becker A, Semmler T, Weiss S, et al. Highly diverse and antimicrobial susceptible Escherichia coli display a naïve bacterial population in fruit bats from the Republic of Congo. PLOS ONE. 2017;12(7):e0178146. doi: 10.1371/journal.pone.0178146.
93. 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. doi: doi:10.1128/mSphere.00333-18.
94. Johnson TJ, Kariyawasam S, Wannemuehler Y, Mangiamele P, Johnson SJ, Doetkott C, et al. The Genome Sequence of Avian Pathogenic Escherichia coli Strain O1:K1:H7

Shares Strong Similarities with Human Extraintestinal Pathogenic E. coli Genomes. Journal of Bacteriology. 2007;189(8):3228-36. doi: doi:10.1128/JB.01726-06.
95. Moulin-Schouleur M, Répérant M, Laurent S, Brée A, Mignon-Grasteau S, Germon P, et al. Extraintestinal Pathogenic Escherichia coli Strains of Avian and Human Origin: Link between Phylogenetic Relationships and Common Virulence Patterns. Journal of Clinical Microbiology. 2007;45(10):3366-76. doi: doi:10.1128/JCM.00037-07.
96. Mora A, López C, Dabhi G, Blanco M, Blanco JE, Alonso MP, et al. Extraintestinal pathogenic Escherichia coliO1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. BMC Microbiology. 2009;9(1):132. doi: 10.1186/1471-2180-9-132.
97. Stephens CM, Skerker JM, Sekhon MS, Arkin AP, Riley LW. Complete Genome Sequences of Four Escherichia coli ST95 Isolates from Bloodstream Infections. Genome Announc. 2015;3(6):e01241-15. doi: 10.1128/genomeA.01241-15.
98. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, Gilbert GL. Distribution of phylogenetic groups, sequence type ST131, and virulence-associated traits among Escherichia coli isolates from men with pyelonephritis or cystitis and healthy controls. Clinical Microbiology and Infection. 2013;19(4):E173-E80. doi: 10.1111/1469-0691.12123.
99. Sora VM, Meroni G, Martino PA, Soggiu A, Bonizzi L, Zecconi A. Extraintestinal Pathogenic Escherichia coli: Virulence Factors and Antibiotic Resistance. Pathogens. 2021;10(11):1355. doi: 10.3390/pathogens10111355.
100. Massot M, Daubié AS, Clermont O, Jauréguy F, Couffignal C, Dahbi G, et al. Phylogenetic, virulence and antibiotic resistance characteristics of commensal strain populations of Escherichia coli from community subjects in the Paris area in 2010 and evolution over 30 years. Microbiology (Reading). 2016;162(4):642-50. doi: 10.1099/mic.0.000242.
101. Marin J, Clermont O, Royer G, Mercier-Darty M, Decousser JW, Tenaillon O, et al. The Population Genomics of Increased Virulence and Antibiotic Resistance in Human Commensal Escherichia coli over 30 Years in France. Appl Environ Microbiol. 2022;88(15):e0066422. doi: 10.1128/aem.00664-22.
102. Bonten M, Johnson JR, van den Biggelaar AHJ, Georgalis L, Geurtsen J, de Palacios PI, et al. Epidemiology of Escherichia coli Bacteremia: A Systematic Literature Review. Clinical Infectious Diseases. 2020;72(7):1211-9. doi: 10.1093/cid/ciaa210.
103. Messerer M, Fischer W, Schubert S. Investigation of horizontal gene transfer of pathogenicity islands in Escherichia coli using next-generation sequencing. PLOS ONE. 2017;12(7):e0179880. doi: 10.1371/journal.pone.0179880.
104. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, et al. WGS accurately predicts antimicrobial resistance in Escherichia coli. J Antimicrob Chemother. 2015;70(10):2763-9. doi: 10.1093/jac/dkv186.

Chapter 5: Genomic analysis of trimethoprim resistant extraintestinal pathogenic Escherichia coli (ExPEC) and recurrent urinary tract infections

5.1. Declaration

Authors:
Dmitriy Li[1], Cameron J. Reid[1], Timothy Kudinha[2], Veronica M. Jarocki[1,*], Steven P. Djordjevic[1,*]

Author affiliations:
[1]ithree institute, University of Technology Sydney, Ultimo, NSW, 2007, Australia
[2]Central West Pathology Laboratory, Charles Sturt University, Orange, NSW, 2800, Australia

Published in:
Microbial Genomics, December 2020, doi: 10.1099/mgen.0.000475

Author contribution statement:
D.L.: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft; C.J.R.: formal analysis, methodology, software; T.K.: data curation, resources; V.M.J.: investigation, methodology, project administration, supervision, validation, writing – original draft; S.P.D.: conceptualization, funding acquisition, project administration, supervision. All authors had a role in writing – review and editing.

5.2. Abstract

UTIs are the most common bacterial infections requiring medical attention and a leading justification for antibiotic prescription. Trimethoprim is prescribed empirically for uncomplicated cases. UTIs are primarily caused by ExPEC and ExPEC strains play a central role in disseminating ARGs worldwide. Here, we describe the whole-genome sequences of trimethoprim-resistant ExPEC and/or ExPEC from recurrent UTIs (67 in total) from patients attending a regional Australian hospital from 2006 to 2008. Twenty-three STs were observed, with ST131 predominating (n=19/67; 28% of isolates), then ST69 and ST73 (both 7%). Co-occurrence of trimethoprim-resistance genes with genes conferring resistance to extended-spectrum β-lactams, heavy metals and quaternary ammonium ions was a feature of the ExPEC described here. Seven trimethoprim-resistance genes were identified, most commonly *dfrA17* (38%) and *dfrA12* (18%). An uncommon *dfrB4* variant was also observed. Two $bla_{\text{CTX-M}}$ variants were identified – $bla_{\text{CTX-M-15}}$ (16%) and $bla_{\text{CTX-M-14}}$ (10%). The former was always associated with *dfrA12*, the latter with *dfrA17*, and all $bla_{\text{CTX-M}}$ genes co-occurred with chromate-resistance gene *chrA*. Eighteen class 1 integron structures were characterized, and *chrA* featured in eight structures; *dfrA* genes featured in seventeen. ST131 H30Rx isolates possessed distinct antimicrobial gene profiles comprising *aac(3)-IIa*, *aac(6)-Ib-cr*, *aph(3')-Ia*, *aadA2*, $bla_{\text{CTX-M-15}}$,

bla$_{OXA-1}$ and *dfrA12*. The most common VAGs were *fimH*, *fyuA*, *irp2* and *sitA* (each gene present in 91 % of isolates). Virulence profile clustering showed ST131 H30 isolates carried similar VAGs to ST73, ST405, ST550 and ST1193 isolates. The sole ST131 H27 isolate carried molecular predictors of EAEC/ExPEC hybrid strains (*aatA*, *aggR*, *fyuA*). Seven isolates (10 %) carried VAGs suggesting ColV plasmid carriage. Finally, SNP analysis of serial UTI patients experiencing worsening sequelae demonstrated a high proportion of point mutations in virulence factors.

5.3. Impact Statement

ExPEC that cause UTIs represent a significant disease burden and contribute greatly to the spread of AMR, both in Australia and worldwide. While WGS provides the most precise means to track resistance and virulence mechanics, to date genomic analyses on UTI-associated ExPEC within Australian hospitals has been limited, particularly for those hospitals situated in regional and remote areas. In this study, ExPEC isolates were taken from patients suffering from cystitis, pyelonephritis and urosepsis attending a Western New South Wales Hospital between 2006 and 2008. We used WGS to investigate diversity, virulence, class 1 integron and AMR gene carriage, and SNPs that occurred in multiple isolates derived from the same patients. By doing so, we provide a baseline for future studies tracking the evolution of ExPEC AMR and virulence potential in Australian UTI-associated ExPEC populations, and particularly the evolution of ExPEC resistant to first-line UTI treatment. This work adds to the national picture of ExPEC, and additionally provides hospital-specific information that may inform future policy making and practices

5.4. Data summary

The 67 draft genomes from the WGS described here have been submitted to the NCBI with the accession numbers SAMN14547713 to SAMN14547779, under BioProject PRJNA623470.

5.5. Introduction

UTIs are the most common bacterial infections to require medical attention and incur an estimated economic burden of ~$2 billion per annum (£1.51 billion; $1=£0.76) in the USA alone (1). In the UK, in 2013–2014, the National Health Service spent £434 million in treating 184 000 patients with unplanned admissions relating to UTIs (2). In Australia, UTIs caused 69 823 hospitalizations in 2017–2018, resulting in 234 455 inpatient days and treatment costs as high as $AU6400 (£3500) per patient (3, 4). Additionally, UTIs are the leading rationale behind antibiotic prescription by GPs (5).

The most common aetiological agent of UTIs are ExPEC, responsible for ~75–95 % of cases (6). Epidemiological studies utilizing MLST indicate that certain pandemic STs account for more than

50% of all ExPEC infections worldwide. These ExPEC lineages are ST131, ST69 (also known as 'clonal group A'), ST10, ST405, ST38, ST95, ST73 and ST127 (7, 8).

ExPEC can enter the intestinal tract via contact with poultry, companion animals, the environment and through sexual contact (9). Additionally, travellers are at higher risk of ingesting ExPEC strains (10). The subsequent introduction of ExPEC into the urinary tract via the urethra can occur due to host behaviours, physiological abnormalities, or medical interventions, such as catheterization (6). After contaminating the urethra, ExPEC can ascend to infect the bladder (cystitis), the kidney (pyelonephritis) or enter the bloodstream (urosepsis) with potentially dire outcomes as ExPEC-associated sepsis has a mortality rate of up to 30% (11). Indeed, each year at least 1.7 million adults in the USA develop sepsis, and 1 in 3 patients who die in hospital have sepsis (12, 13).

ExPEC possess an array of VAGs that enable ascension, colonization and persistence within a hostile and nutrient-deficient environment. These VAGs – commonly located on chromosomal PAIs or encoded on virulence plasmids – include uroepithelial adhesins, such as P-fimbriae (*pap* genes), S-fimbriae (*sfa* genes), F1C-fimbriae (*foc* genes), and Dr adhesins (*dra* genes, *afa* genes), toxins including secreted autotransporter toxin (*sat*) and cytotoxic necrotizing factor 1 (*cnf1*), capsule (*kpsMII*), and several iron-acquisition systems, such as aerobactin (*iuc* genes, *iutA*), salmochellin (*iro* genes) and yersiniabactin (*ybt* genes, *irp1*, *irp2*, *fyuA*) (14, 15). ExPEC have also been associated with a greater abundance of bacteriocins, toxins that inhibit the growth of other *E. coli* (16). These bacteriocins are frequently carried on virulence plasmids, such as ColV plasmids (17). While no singular VAG can be attributed to ExPEC pathogenesis, a molecular definition of ExPEC has been proposed as being *E. coli* that contain at least two of the following VAGs: *papA* and/or *papC*, *sfa*/*foc*, *afa*/*draBC*, *kpsM II* and *iutA* (18). Furthermore, the presence of certain VAGs have been associated with specific uro-clinical syndromes, such as P- and S-fimbriae, Dr adhesins, and the invasin *ibeA* with pyelonephritis (6), and yersiniabactin *irp2* and *fyuA* with sepsis (19). *fyuA* is also directly linked to ExPEC biofilm formation in human urine (20), as is the salmochelin receptor *iroN* (21).

In Australia, severe pyelonephritis and sepsis are typically treated with intravenous gentamicin and amoxicillin, while cystitis is typically treated empirically with either nitrofurantoin or trimethoprim (22). However, a five-year study on AMR in urinary *E. coli* from an Australian metropolitan hospital found significant increases in trimethoprim, nitrofurantoin, amoxicillin, and gentamicin resistance, as well as significant extended-spectrum β-lactam resistance (23). ExPEC, particularly ST131 strains, have been central to a worldwide increase in ESBL-producing

E. coli. The ST131 lineage consists of three major clades, each of which is strongly associated with specific *fimH* alleles. Strains from clade C typically carry *fimH30* and are currently the most prominent subtype identified in human infections (24). Clade C can be further subdivided into two distinct subclades, namely C1 and C2. Subclade C1 primarily consists of strains exhibiting fluoroquinolone resistance due to *gyrA* and *parC* SNPs. These *fimH30* resistance strains are designated H30R. On the other hand, subclade C2 comprises strains that possess both fluoroquinolone resistance SNPs and plasmid-associated $bla_{CTX\text{-}M}$ genes, thus leading to their designation as H30Rx. (25–27). *E. coli* ST131 first emerged in the literature in the early 2000s and has been linked to a 300 % increase in USA hospital admissions due to ESBL-producing *E. coli* in the following decade (28, 29). The global rapid increase in ESBL-producing *E. coli* has had serious knock-on effects, driving carbapenem prescription and in turn promoting the spread of potentially untreatable carbapenemase-producing *E. coli* (30).

Like all *E. coli*, ExPEC have highly flexible genomes and a proclivity to capture and disseminate genes through HGT (31). HGT can occur via plasmid conjugation, phage transduction and via small MGEs, such as transposons, IS and integrons, allowing for inter- and intra-species movement of genetic information (32). As observed with the global dissemination of $bla_{CTX\text{-}M}$, a single MGE capture event can have worldwide repercussions (30).

AMR surveillance of ExPEC populations within high selective pressure environments, such as hospitals, could provide meaningful insights and inform future policy making and practices. In addition to the aforementioned five-year study on AMR within ExPEC isolated in an Australian hospital, a recent Australian government report also stated that *E. coli* resistances to ESBLs and fluoroquinolones are climbing (4). While these reports highlight national trends, they tend to focus on hospitals situated in metropolitan areas, while more remote and rural areas get neglected. Furthermore, genomic surveillance of ARGs and VAGs within Australian hospital *E. coli* populations is currently limited.

Here, we used WGS to characterize 67 trimethoprim resistant ExPEC strains from patients with UTIs collected in a large Western NSW hospital catering to regional, remote and rural Australian communities. The genomes were typed using Clermont phylogrouping, e-serotyping and MLST, and screened for the presence of ARGs, VAGs, PAIs and plasmid replicons, and several AMR associated class 1 and class 2 integrons found in this collection were characterized. Furthermore, isolates from recurrent infections, including those with a subsequently more severe clinical sequelae, were interrogated for SNPs leading to point mutations.

5.6. Methods

5.6.1. Sample collection and selection criteria for WGS

Urine specimens were collected by clinical staff members of participating health-care centres using a standardized protocol. Semi-quantitative culture was performed on horse blood, MacConkey and chromogenic agars, followed by conventional identification. Isolates were stored in 50% (v/v) glycerol in trypticase soy broth at −70 °C.

5.6.2. Phenotypic resistance testing

The isolates were tested for susceptibility to 14 antibiotics as per the disc diffusion method specified by the Clinical and Laboratory Standards Institute (33), using Neo-Sensitabs discs (Rosco). The antibiotics tested were (disc content) amikacin (30 µg), amoxicillin–clavulanate (60 µg), ampicillin (25 µg), ceftazidime (30 µg), ceftriaxone (30 µg), cephalothin (30 µg), ciprofloxacin (10 µg), gentamicin (10 µg), imipenem (10 µg), nalidixic acid (30 µg), nitrofurantoin (300 µg), norfloxacin (10 µg), tetracycline (30 µg) and trimethoprim–sulfamethoxazole (5 µg). The double-disc diffusion test was used to detect the production of ESBLs (34).

5.6.3. DNA isolation and WGS

DNA extraction and WGS were performed as described previously (35). Briefly, DNA was extracted using the ISOLATE II genomic DNA kit (Bioline) and stored at −20 °C. Short-read sequencing was performed using an Illumina HiSeq 2500 v4 sequencer in rapid PE150 mode.

5.6.4. Genome assemblies

Raw reads were used to assemble draft genome sequences via Shovill software using default settings (https://github.com/tseemann/shovill). Assemblies underwent quality control using assembly-stats software (https://github.com/sanger-pathogens/assembly-stats). Assembly statistics for this collection are available in Appendix 8.

5.6.5. Phylogenetic analysis

A maximum-likelihood phylogenetic tree of this collection was reconstructed using IQ-TREE 2 (www.iqtree.org/) (36) and a SNP-based phylogenetic tree was reconstructed using snplord (github.com/maxlcummins/pipelord/snplord/tree/master/snplord), an automated pipeline that utilizes snippy (github.com/tseemann/snippy), Gubbins (37) (github.com/sanger-pathogens/gubbins) and SNP-sites (github. com/sanger-pathogens/snp-sites). The SNP-based tree of the entire collection was built using 76.0 % (3,525,942 bp/4,639,675 bp) of the reference genome K12 MG1655 (GCA_000005845.2) and consisted of 35,038 SNPs (141,270 SNPs before recombination filtering). Trimethoprim-sensitive ExPEC isolates ERR434278, ERR434751, ERR434270, ERR434273 and ERR434271 (38), and trimethoprim-sensitive EAEC isolates SRR5470250, SRR5024242 and SRR3574247 (39), all representing the major STs in this collection,

were added to both the phylogenetic analysis and gene screening as controls. An additional SNP-based ST131 only tree was built using 92.03% (4,831,235 bp/5,249,449 bp) of reference genome ST131 EC958 (GCA_000285655.3) to identify SNP sites. Recombination filtering reduced 7,460 SNPs to 386 SNPs. The ST131 SNP-based tree was reconstructed using IQ-TREE 2 and a recombination filtered alignment of 478 bp produced by Gubbins. All trees were visualized using the iTOL v4 online-based software (40) (itol.embl.de/). The ExPEC pangenome was calculated using Roary v3.11.2 (41) and visualized using Phandango (42). Clonal samples from serial patients were removed from this analysis (total of 57 isolates used) to prevent deflation of the core genome.

5.6.6. Serial patient SNP analysis

The snplord pipeline was run on isolates originating from a single patient using the isolate with the earliest isolation date as a reference in each instance. SNPs called by this pipeline were checked manually in .gbk files generated by Prokka (43) using SnapGene. Gubbins output was used to determine which SNPs were the result of homologous recombination.

5.6.7. Gene screening

Isolate STs, serogroups and phylogroups were determined in silico using MLST v2.0 (44) (cge.cbs.dtu.dk/services/MLST), SerotypeFinder v2.0 (45) (cge.cbs.dtu.dk/services/SerotypeFinder) and ClermontTyping (46) (enterobase.warwick.ac.uk/), respectively. ST131 isolates additionally underwent *fimH* typing using FimTyper (47) (cge.cbs.dtu.dk/services/FimTyper/). The ARIBA read-mapping tool (48) was used the screen for ARGs, plasmid replicons and VAGs using the following reference databases: ResFinder (49) (cge.cbs.dtu.dk/services/ResFinder/), PlasmidFinder (50) (cge.cbs.dtu.dk/services/PlasmidFinder/) and VirulenceFinder (51) (cge.cbs.dtu.dk/services/VirulenceFinder/). However, as VirulenceFinder screens for only one *papA* allele, an additional blastn search was also performed using 14 *papA* alleles (52). An additional custom database with AMR-associated ISs, class 1 and 2 integrases, and additional ExPEC-associated VAGs was also utilized and can be accessed at github.com/CJREID/custom_DBs. ARIBA data were processed using a bespoke script accessible at github.com/maxlcummins/pipelord/tree/master/aribalord and visualized using the following R packages: ggplot2 v3.3.0 (github.com/tidyverse/ggplot2) and ggtree v2.2.1 (53) (github.com/YuLab-SMU/ggtree).

To infer ColV plasmid and PAI carriage, short reads from each isolate were mapped to the ColV reference plasmid pCERC4 (KU578032), PAI-I_{CFT073} (AE014075.1, start 3,406,225 bp, end 3,469,205 bp), PAI-II_{536} (AJ494981), PAI-III_{536} (AF301153) and PAI-IV_{526}/HPI using the BWA v0.7.17 (54) and converted to BAM file format using SAMtools v0.1.18 (55). A bespoke script,

available at github.com/maxlcummins/pipelord/tree/master/plasmidlord, was used to produce a histogram of read depth as a function of reference coordinate and clustered based on their Euclidean distances, and used to generate a heatmap.

BLASTn (56) was used to determine whether integrons characterized in this collection are present in other genomes available in public databases, and associated metadata were pulled from GenBank (56) and Enterobase (57). Integron B structure comparison was achieved using EasyFig (58). All gene schematics were generated using SnapGene (snapgene.com/).

5.6.8. Statistical analysis

R software v4.0.2 was used for the reconstruction of correlation heat maps and classical (metric) MDS analysis, using the ggplot2 v3.3.0 (ggplot2.tidyverse.org/) package for visualization, and the ggcorrplot v0.1.3 (github.com/kassambara/ggcorrplot) package for correlation calculation. Two MDS analyses were performed, one for identified virulence genes and one for AMR genes, using the standard R package functions dist and cmdscale in conjunction with a presence/absence matrix (with one for presence and zero for absence).

5.7. Results and Discussion

Genomic surveillance using WGS can provide high-resolution data on ARG presence, persistence, evolution, and potential for horizontal transfer (59), as well as inform on population diversity and VAG carriage, and provide in-depth analyses of SNP-mediated mutations. Yet to date, genomic analyses on UTI-associated ExPEC within Australian hospitals has been limited. Here, we provide a retrospective WGS study on ExPEC resistant to first-line trimethoprim treatment and/or isolates collected from serial UTIs between 2006 to 2008 from a regional hospital in Western NSW, Australia. We aimed to provide a baseline for future studies tracking the evolution of ExPEC AMR and virulence potential in Australian UTI-associated ExPEC populations by reporting on ARG, VAG and MGE carriage, characterizing 18 class 1 integrons structures identified in this collection and identifying SNPs occurring in recurrent UTI isolates.

5.7.1. Demographic and clinical characteristics

From clinical samples collected between 2006 and 2008, 76 isolates were selected for genome sequencing due to phenotypic trimethoprim resistance and/or originating from a serial UTI patient. After quality-control measures, the final collection comprised 67 *E. coli* draft genomes (mean contig size of 23,477 bp, with a mean number of contigs of 234) originating from 37 samples taken from patients with cystitis (55%), 23 samples from patients with pyelonephritis (34%) and 7 samples from patients with urosepsis (10%) (Figure 5.1a). Recurrent UTIs are common, particularly in women, with 27% reporting a recurrence within 6 months (53). In this

collection, 12 out of 51 patients were sampled more than once due to multiple, or ongoing, UTIs during the study period.

Women are more likely to contract UTIs than men at a ratio of 8:1, and one in three women will experience at least one UTI necessitating antibiotics by age 24 (60). The risk of UTIs tends to increase with age in both sexes (61); however, the prevalence of UTIs in women over the age of 65 is double that of the overall population (62). This collection is reflective of UTI epidemiology with more isolates derived from female patients (44 vs 7; 86 vs 14%) and with patients' ages ranging from 13 months to 91 years, the majority (64%) being above 66 years. The mean age for female patients was 66 years, and 74 years for male patients (Figure 5.1b). No patient was identified as pregnant.

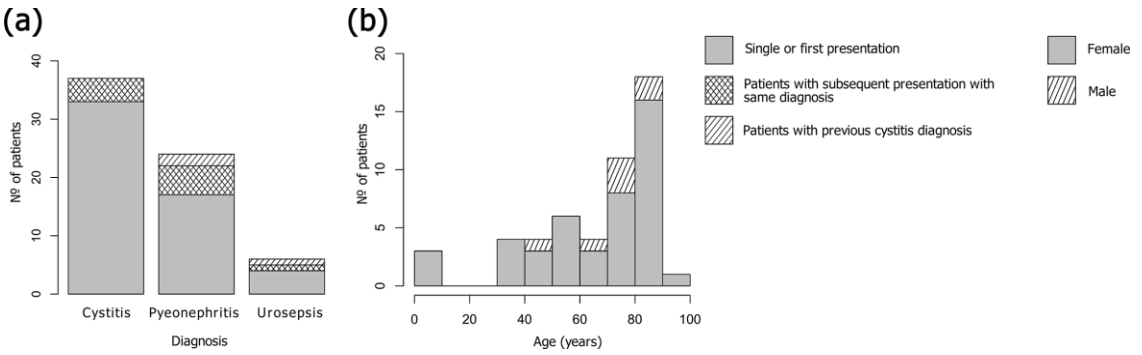


Figure 5.1: Sample population and clinical characteristics. a) Diagnoses of patients presenting with UTIs over the study period. For patients with multiple samples taken for the same diagnosis (horizontal stripes): For cystitis, the average days between collections was 126 days, for pyelonephritis 63 days, and 68 days for urosepsis. For patients with a previous cystitis diagnosis and subsequent pyelonephritis or urosepsis diagnoses (sloped stripes) the average length between collections was 174 days, and 181 days urosepsis (n = 1). **b)** Age and biological sex of patients. Females represented in grey, males as shaded.

5.7.2. Phylogenetic analysis

While ExPEC are phylogenetically diverse, a recent systematic review and meta-analysis of 217 studies flagged a handful of globally dominant STs, the most prominent being ST131 (phylogroup B2), followed by ST69 (D), ST10 (A), ST405 (D), ST38 (D), ST95 (B2), ST648 (D) and ST73 (B2) (7). Consequently, B2 and D constitute the phylogroups most frequently associated with ExPEC infections. To ascertain the evolutionary relationships between samples in this collection, a maximum-likelihood phylogenetic tree was reconstructed using the complete genome of *E. coli* strain K12-MG1655 as a reference (Figure 5.2a). The tree demonstrates clear clustering based on phylogroup and ST (Figure 5.2a). The collection comprised 23 STs, and despite selecting for trimethoprim resistance, the collection followed global trends with ST131 (O25:H4) being the most common (*n* = 19; 28%), followed by ST69 (ONT:H18 and O11:H18) and ST73 (O22:H1,

O25:H1, O6:H1, and ONT:H1) with five isolates each. Isolates from phylogroup B2 were the most prevalent (*n* = 38; 56 %), and the least prevalent were from B1 (*n* = 3; 4 %).

The fact that ST131 isolates, which are associated with ESBL resistance, outnumbered ST69 isolates, which are associated with trimethoprim resistance (63), is reflective of the high number of UTIs caused by ST131, particularly in association with pyelonephritis, within this regional NSW community (64–66). To resolve the ST131 isolates into clades, we screened for *fimH* alleles in conjunction with a separate SNP-based phylogenetic tree (Figure 5.2b). Only one ST131 isolate, HOS40, carried H27; thus, it belonged to clade B, a clade associated with animal to human transmission (67). All other ST131 isolates carried H30, thereby belonging to the globally dominant clade C (24). Of the ST131 H30 isolates, 7 (37 %) were H30R (clade C1) and 11 (61 %) were ESBL $bla_{\text{CTX-M-15}}$ -associated H30Rx (clade C2) (25). SNP analyses showed that the H27 isolate differed from H30 by a mean of 277 SNPs, and that H30R isolates differed from H30Rx by a mean of 93 SNPs. The H30Rx isolates originated from different patients, situated in seven different postcodes (~300 km between the two most distant postcodes), caused either cystitis or pyelonephritis, and were isolated over a period of 569 days. Yet despite these differences, the H30Rx isolates differed by only seven SNPs on average (across 99.6 % of the reference H30Rx HOS54 genome), indicating a persistent clone in this community. Conversely, H30R isolates differed by a mean of 58 SNPs, due to the presence of two distinct branches (Figure 5.2b; i and ii), and the SNP difference between H30R and H30Rx branches was 140. The pangenome for this collection consisted of 14 091 genes making up a core genome (present in 100% of isolates) of 2886 genes and an accessory genome of 11 205 genes, leading to a total pangenome length equal to 11.32 Mb, and can be viewed in Appendix 9.

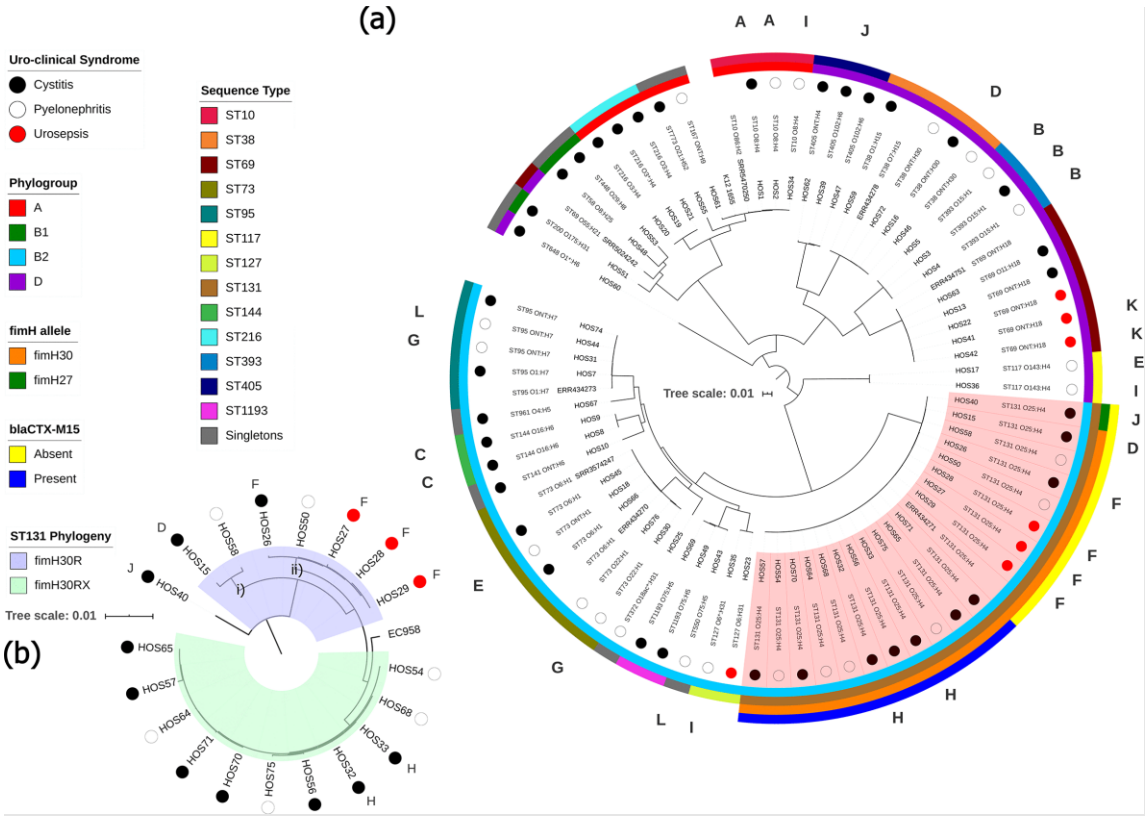


Figure 5.2: Maximum-likelihood phylogenetic trees showing genetic relatedness of ExPEC strains. Tree scale bars represent number of substitutions per site of alignment. **a)** Mid-point rooted maximum-likelihood phylogenetic tree, inferred using IQ-TREE 2 and K12-MG1655 as a reference, containing the 67 ExPEC isolates sequenced in this study. Coloured circles represent uroclinical syndrome (black, cystitis; white, pyelonephritis; red, urosepsis). The inner-most ring represents the phylogroup (A, red; B1, green; B2, light blue; D, purple). The next ring represents ST. The two outermost rings apply to ST131 isolates only (marked in red shaded area), the inner ring highlights *fimH* alleles (orange, fimH30; green, H27) and outer ring shows the presence or absence of $bla_{\text{CTX-M-15}}$ (blue, present; yellow, absent). Letters around the perimeter represent multiple isolates taken from a single patient, one letter per patient. **b)** SNP-based phylogenetic tree, inferred using IQ-TREE 2, resolving ST131 isolates into clades. H30R, blue shaded area; H30Rx, green shaded area. Trimethoprim-susceptible ExPEC isolates ERR434278, ERR434751, ERR434270, ERR434273 and ERR434271, and trimethoprim-susceptible EAEC isolates SRR5470250, SRR5024242 and SRR3574247 were added as controls.

5.7.3. Phenotypic resistance

ExPEC have been a driving force behind a worldwide increase in ESBL-producing *Enterobacteriaceae* (68). Furthermore, ExPEC isolates have been reported as carrying transmittable resistances to carbapenem and colistin (69, 70), both being last-line antibiotics for MDR Gram-negative bacteria. The WHO has flagged increasing cases of MDR bacteria worldwide as one of the most serious public-health threats and deemed surveillance as a strategic priority of the Global Action Plan on AMR (71). Despite trimethoprim remaining a first-line empirical treatment option for UTIs, resistance in uropathogens is increasing worldwide. In Australia, resistance to trimethoprim in *E. coli* has increased from 16.6 % in 2004 (72) to 31.2 % in 2017

(73). These figures and trends are similar to those in other developed countries, though lower than in developing countries were trimethoprim resistance in UTI-associated ExPEC ranges from 54% to 82% (74). Isolates were tested against 11 antibiotics belonging to 8 classes (Figure 5.3). The mean number of antibiotics an isolate was resistant to was five, with two isolates (HOS56 and HOS70, both ST131 H30Rx) being resistant to 10/11 tested (all bar the carbapenem imipenem). The most-common resistance phenotype, aside from trimethoprim and trimethoprim/sulphonamide, was ampicillin (*n* = 50; 75%). High levels of ampicillin resistance in UTIs have been described worldwide, leading to governing bodies recommending its disuse (4). The least common phenotypic resistances were to imipenem (*n* = 2; 3%) and the anaerobic DNA inhibitor nitrofurantoin (*n* = 4; 6%). This hopeful observation indicated that both a first and a last-line treatment option for UTIs remained largely efficacious, at least in the past, though low levels of nitrofurantoin and imipenem resistance continue to be reported in UTI isolates currently (75, 76). Antibiotic resistance varied by ST (Figure 5.3a) and uro-clinical syndrome. Isolates originating from patients with cystitis and pyelonephritis shared similar phenotypic resistance profiles; however, isolates from urosepsis patients tended to be more sensitive and no isolate was resistant to the aminoglycoside gentamicin, the cephalosporin cefotaxime, nitrofurantoin or imipenem (Figure 5.3b). A high co-occurrence (>30%) of trimethoprim and ESBL resistance was recently reported in the USA (77). Furthermore, Mulder et al. (78) demonstrated that previous use of extended-spectrum β-lactams in patients with UTIs was significantly associated with trimethoprim resistance. These data are reflective of treatment practices and suggests a stepwise acquisition of resistances to antimicrobials over time, having become resistant to first-line treatment and then increasingly to ESBLs.

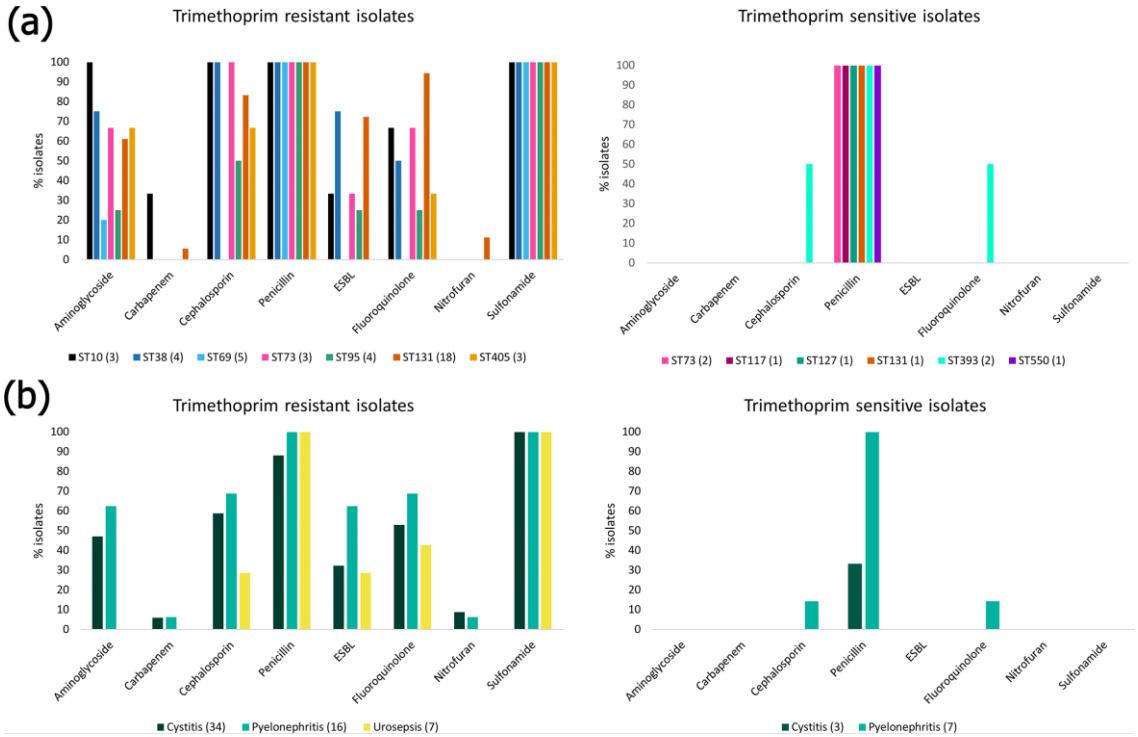


Figure 5.3: Phenotypic resistance of ExPEC isolates. a) Phenotypic resistance profiles by ST. For trimethoprim-resistant isolates, only STs with three or more isolates were included. Full phenotypic resistance profiles can be viewed in Appendix 10. **b)** Resistance phenotypes by uro-clinical syndrome. Following antibiotics represent antibiotic class: Aminoglycoside - Gentamicin, Carbapenem - Imipenem, Cephalosporin - Cefotaxime and Cefalexin, Penicillin - Ampicillin, ESBL - ESBL producing, Fluoroquinolone - Ciprofloxacin and Norfloxacin, Nitrofuran - Nitrofurantoin, Sulfonamide - Sulfamethoxazole.

5.7.4. Genetic screening

This collection was screened for ARGs, ISs associated with AMR, class 1 and 2 integrases (*intI1*, *intI2*), heavy-metal-resistance genes, VAGs and plasmid replicons using ARIBA (Figure 5.4). A full list of identified genes and associated metadata for each isolate is available in Appendix 10.

5.7.4.1. ARGs

AMR in uropathogens complicates the treatment of UTIs. An advantage of WGS in AMR surveillance is the level of detail and precision gleaned from identifying specific ARG alleles, and co-resistances not tested on standard antibiotic panels, including disinfectant and heavy-metal-resistance genes (59). Here, we identified a total of 40 ARGs (Figure 5.4), with the majority of isolates (*n* = 56; 84 %) carrying at least three ARGs.

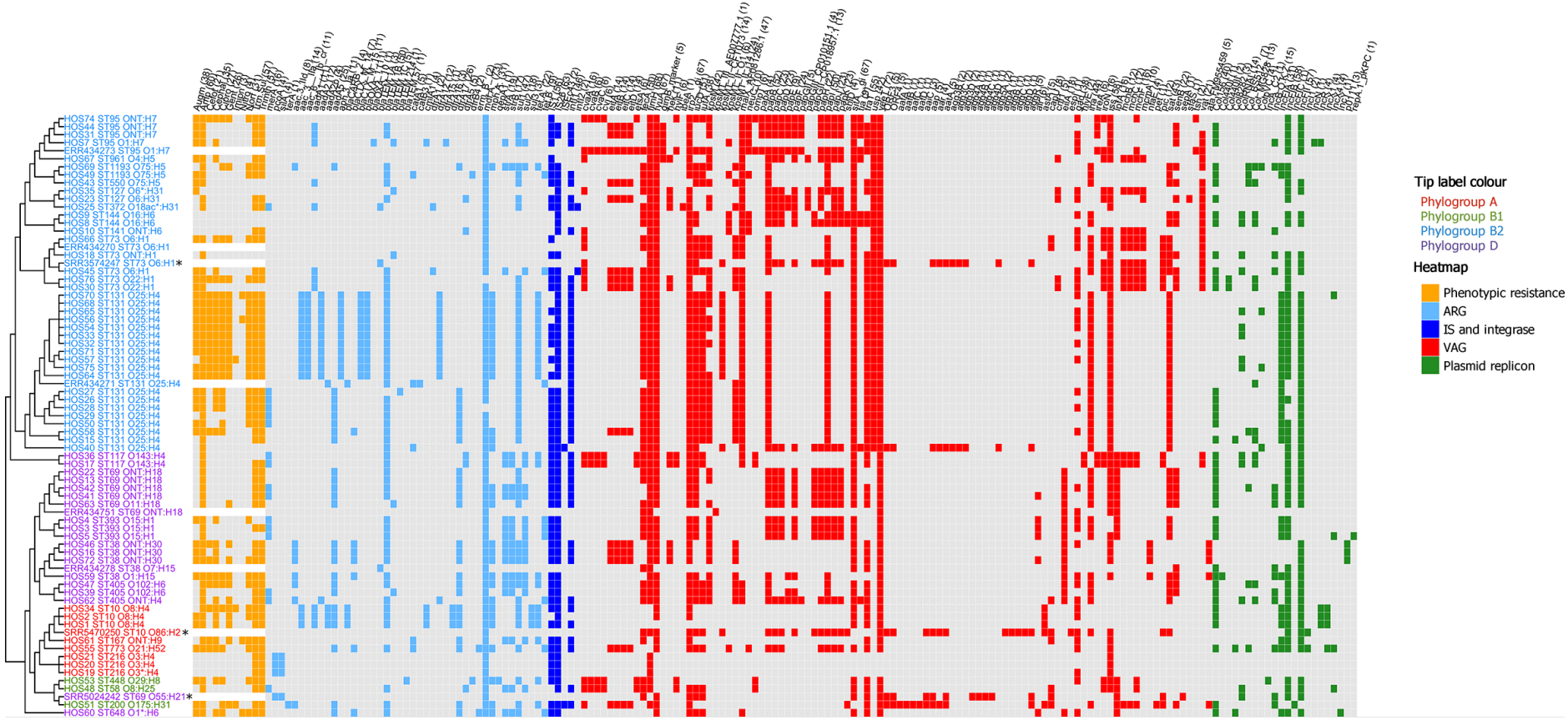


Figure 5.4: Genotypic profile of 73 ExPEC isolates. Tip label colour corresponds to phylogroups (red, A; green, B1; teal, B2; purple, D). The adjacent heatmap represents phenotypic resistances and intact genes present in each isolate (orange, phenotypic resistance; light blue, ARG; dark blue, IS and class 1 and 2 integrase (*intI1*, *intI2*); red, VAG; green, plasmid replicon). Along the top of the heatmap, the numbers in brackets after gene names represent the total number of isolates with trait present). Trimethoprim-susceptible ExPEC isolates ERR434278, ERR434751, ERR434270, ERR434273 and ERR434271 and trimethoprim-susceptible EAEC isolates SRR5470250, SRR5024242 and SRR3574247 (marked by asterisks) from other studies were added as controls. The tree on the left side is the maximum-likelihood tree seen in Figure 5.2 (reference K12 MG1655 omitted) presented in cladogram form.

The most common mechanism for acquiring trimethoprim resistance is through the acquisition of *dfr* genes (79). There are currently more than 40 identified *dfr* variants (80). These are often associated with MGEs, such as plasmids and transposons, and are almost exclusively observed as resistance gene cassettes within the variable regions of class 1 and 2 integrons in human (81), animal (35, 82, 83) and environmental (84) *E. coli* isolates, resulting in their rapid spread. Seven trimethoprim-resistance genes were identified in this collection, the most common being *dfrA17* (*n*=26; 38%), followed by *dfrA12* (*n*=12; 18%). Several ESBL genes were also detected, the most common being of the $bla_{\text{CTX-M}}$ type (*n* = 18, 26%). $bla_{\text{CTX-M}}$ progenitor genes are thought to have originated from the chromosomes of soil bacteria (85) and their subsequent capture by predominately IncF family plasmids has led to their global dissemination in both clinical and non-clinical settings (30). IncF family replicons were the most prevalent replicons detected in this collection, the foremost being IncFIB (*n* = 53; 79%), followed by IncFII (*n* = 51; 76%). Additionally, IncFIA replicons were detected in 30 isolates (44%). While $bla_{\text{CTX-M}}$ ESBLs comprise a heterogeneous family of enzymes, the $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-14}}$ variants are currently most prevalent worldwide and are strongly associated with ST131 (30), though other STs have also contributed to their spread (86). In this historic collection, $bla_{\text{CTX-M-15}}$ (*n* = 11; 16%) and $bla_{\text{CTX-M-14}}$ (*n* = 7; 10%) were the only two $bla_{\text{CTX-M}}$ variants identified. All $bla_{\text{CTX-M-15}}$ genes were carried solely by ST131 H30Rx isolates (Figure 5.2a). Conversely, the $bla_{\text{CTX- M-14}}$ genes were present in four ST38 isolates, the sole examples of ST648 and ST773 and an H30R ST131 isolate. UTI isolates carrying $bla_{\text{CTX-M}}$ genes have previously been noted exhibiting high co-resistance to trimethoprim (77). In this collection, all isolates carrying $bla_{\text{CTX-M-15}}$ also carried *dfrA12*, and all isolates carrying $bla_{\text{CTX-M-14}}$ co-occurred with *dfrA17*. Three additional ESBL genes were identified, $bla_{\text{OXA-10}}$, $bla_{\text{TEM-214}}$ (87) and $bla_{\text{TEM-57}}$ (88), occurring in one ST95 isolate, one ST131 H30R isolate and one ST448 isolate, respectively.

In addition to trimethoprim- and ESBL-resistance genes, other common ARGs identified were NSBL gene $bla_{\text{TEM-1B}}$ (*n* = 49; 73%), sulphonamide-resistance gene *sul1* (*n* = 46; 69%), aminoglycoside-resistance gene *aadA5* (*n* = 25; 37%), tetracycline-resistance gene *tetA* (*n* = 22; 30%) and macrolide-resistance gene *mphA* (*n* = 31; 46%). Fluoroquinolone-resistance genes *aac(6)-Ib-cr* (*n* = 11; 16%) and *qepA1* (*n* = 1; ST448) were also detected. Isolates belonging to phylogroup D contained the most ARGs on average (nine per isolate), consistent with reports stating that an MDR profile in ExPEC is most associated with this phylogroup (89–91). However, three ST10 isolates of phylogroup A carried the highest number of ARGs overall (*n* = 14). An MDS analysis demonstrated that ST131 H30Rx isolates possessed the most distinct ARG profiles (Figure 5.5, red area), which could be attributed to the strong correlation of *aac(3)-IIa*, *aac(6)-*

Ib-cr, *aph(3')-Ia*, *aadA2*, *bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$ and *dfrA12* in this ST and clade (Appendix 11). Notably, all the trimethoprim-sensitive ExPEC and EAEC isolates used as controls in this study clustered together with four trimethoprim-sensitive isolates from this collection (Figure 5.5, blue area), with the exception of one control ExPEC ST131 isolate (Figure 5.5, red arrow).

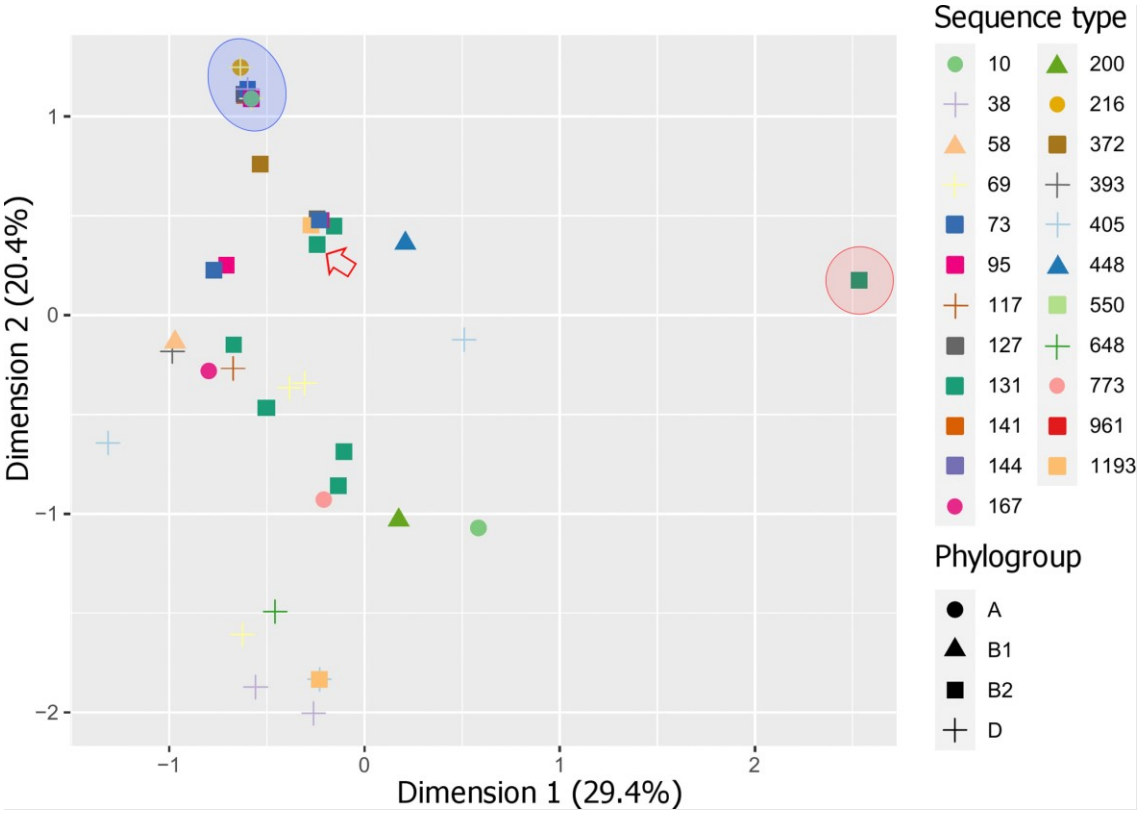


Figure 5.5: MDS analysis of ARGs detected in the ExPEC collection. Red area, ST131 H30Rx cluster (11 isolates share the same coordinate); blue area, trimethoprim-sensitive control isolates clustered with four trimethoprim-sensitive isolates from this collection, with the exception of one control ST131 isolate (red arrow). The percentage of total explained variation for each dimension is indicated in parentheses after each axis label.

Heavy-metal- and detergent/disinfectant-resistance genes are often carried by the same MGEs harbouring ARGs, sparking concern and growing evidence that selective pressures induced by their extensive use in industry, agriculture, and health-care facilities, as well as via contamination, can co-select for AMR (92–94). The hospital wherein our isolates originated caters for a catchment area significantly associated with agriculture, including food animals (bovine, sheep, pig and poultry), and various crops, such as fruits and nuts (95). Thus, the isolates were screened, and two detergent-resistance genes – *qacEΔ1* (*n* = 47; 70%) and *qacH* (*n* = 3; 4%) – conferring resistance to quaternary ammonium compounds, and six heavy-metal-resistance genes were identified. All isolates carried *zntA*, which confers resistance to zinc, lead and cadmium. We also identified *chrA* (*n* = 32; 48%), corresponding to chromium resistance, and *merA* (*n* = 16; 24%), which confers mercury resistance. Silver- and copper-resistance genes

silA and *pcoA* were present in only three isolates (4%; all ST216) and tellurium-resistance gene *terA* was observed in a single ST200. Heavy-metal-resistance genes, including *zntA*, *merA*, *pcoA* and *chrA*, are known to be significantly associated with detergent resistance genes, and *sul*, *tet*, bla_{TEM}, bla_{SHV} and bla_{CTX} variants (93). Furthermore, chromium resistance has been positively correlated with relative $bla_{CTX\text{-}M}$ and bla_{OXA} gene abundance (96). Here, 100% of isolates carrying $bla_{CTX\text{-}M}$ genes also carried *chrA*, and 91% of isolates with bla_{OXA} also harboured *chrA*.

5.7.4.2. Class 1 and 2 integrons

The presence of class 1 or 2 integrons is considered a reliable proxy for an MDR genotype (97). Class 1 integrons are more prevalent than class 2 integrons (97) and typically contain two conserved regions – 5'-CS and 3'-CS – flanking a variable gene cassette. The *intI1* integrase gene, responsible for capturing, removing and rearranging genes within the variable cassette, is located in the 5'-CS region. The 3'-CS region typically contains a truncated but functional *qacEΔ1* gene fused to a *sul1* gene (98). Class 1 and 2 integrons are mobilized and spread by MGEs, such as transposons and conjugative plasmids. Additionally, the insertion element IS*26* is renowned for its ability to capture and assemble ARGs in CRRs (32) and alter the structure of class 1 integrons (35, 98). Here, IS*26* was identified in 63 (94%) isolates, including all 45 (67%) that carried *intI1*.

Class 1 and 2 integron loci are often rich in repetitive sequences, posing challenges to assembling complete structures using short-read sequencing data. Despite this limitation, we resolved 18 class 1 integron structures, including ARGs and MGEs found downstream of the typical 3'-CS region (Figure 5.6a) and 1 class 2 integron (Figure 5.6b). BLASTn and GenBank were used to determine whether these integrons had been previously deposited into public databases (Table 5.1). In general, the integrons from this collection are also found in ExPEC-associated STs, most commonly ST131 and ST617 (of ST10 CC). They are most frequently seen on IncF plasmids of varying pMLSTs, in a range of hosts, including humans, dogs and gulls, and from a range of geographical locations, including Europe, Asia, North America and Australia (Table 5.1). The most common genes located in the variable gene cassette were *dfrA17* and *aadA5* (Figure 5.6a).

Integron k was the most common structure in this collection, present in 12 isolates (from ST131 and ST405), and carried *dfrA12* (trimethoprim), *aadA2* (aminoglycoside), *qacEΔ1* (quaternary ammonium compounds), *sul1* (sulphonamide) and *mphA* (macrolide) resistance genes. The same integron was described in an ST131 H30Rx isolate taken from a French patient a year after contracting pyelonephritis while travelling in Nepal (99). The integron k structure also carried the aforementioned chromium resistance gene *chrA*, which in turn was also present in seven

other integron structures (d, e, f, g, l, m and r). Class 1 integrons are often translocated via mercury-resistance transposons carrying genes of the *mer* operon (81), and such transposons were present in integrons i, j and n structures. Integrons n and r also carried a recently characterized *dfrB4* gene conferring high-level resistance to trimethoprim. Both isolates that carry these integrons (ST1193 and ST448) predate the first report of this gene in a clinical sample (UTI) in 2017 (100). Additionally, integron r carried an uncommon plasmid-mediated fluoroquinolone-resistance gene, *qepA*. Integron b was the only structure to carry an ESBL gene (bla_{OXA-10}) and carried additional genotypic resistance genes for trimethoprim (*dfrA15*), chloramphenicol (*cmlA1*), aminoglycosides (*aadA1*) and sulphonamides (*sul1*). An identical structure to integron b could not be found in public databases; however, a similar structure was observed chromosomally in *Pseudomonas aeruginosa* strain SP4527 (ST357) isolated from a patient with bacteraemia in India, in 2016 (Figure 5.6b). Integron c consists of a short *dfrA5*-IS26 configuration that has been detected in, and is identical to, *E. coli* strains sourced from cattle located on several NSW properties (98, 101), and in commensal *E. coli* from Australian pigs (35). Furthermore, the signature has been used to track ColV plasmids carrying a complex resistance locus in patients with UTI and urosepsis in Australia in *E. coli* ST58 (102, 103). These studies highlight how plasmids that play an important role in carrying intestinal pathogenic *E. coli* virulence genes (101) and extraintestinal pathogenic virulence gene cargo (102, 103) can be tracked by the presence of this unique signatures. Moreover, it shows that transposon belonging to the Tn*3* family of mercury-resistance transposons can mobilize complex resistance loci on diverse plasmid backbones. Integron l has been described within a larger salmonella genomic island 1 (SGI1) structure in *Proteus mirabilis* (104). The integron a structure carried the sole example of a *sul3* gene, which has been associated with Australian *E. coli* isolates from pig farms (35, 83). This integron, found in our ST10 isolates, also uniquely carried *dfrA16*, bla_{CARB-2} and *qacH*. Integron a has been located on IncFIA and IncI plasmids in CC10 (ST617) *E. coli* isolates originating from both human clinical samples from Germany and the USA, a broiler chicken sample from Belgium (Table 5.1), as well as farmed red deer from Spain (105). Incidentally, the patients from whom these isolates originated resided in proximity to two red deer farms. Lastly, the class 2 integrase *intI2* gene was identified in one ST372 and one ST73 isolate (Figure 5.6b). The *intI2* from this class of integron is inactive, impeding its ability to modify gene cassettes. Thus, class 2 integron cassettes are highly conserved (97). The *dfrA1-sat2-aadA1* cassette found in this collection represents the most common array and is particularly prevalent in UTI-associated *Proteus* species (106).

Table 5.1: Summary of BLAST hits to integrons found in this collection. Rows in blue indicate integrons with no 100% matches. The closest integron match is provided. The closest integron match is provided. NA - Not Applicable. * Only 100% matches from *E. coli* are shown.

Integron	Accession no.	% Coverage; % Identity	Integron location	Inc group (pMLST)	Isolation Date	Species (ST) / (STs from this collection)	Host	Host Disease	Host location
a	MK070495.1	100%; 100%	Plasmid	I1 (CC-3)	2013	*E. coli* (NA) / (ST10)	Chicken	NA	Belgium
	CP018105.1	100%; 100%	Plasmid	F (F-:A18:B-)	2016	*E. coli* (ST617)	Human	NA	Germany
b	CP034409.1	93%; 100%	Chromosome	-	2016	*P. aeruginosa* (ST357) / (ST95)	Human	Sepsis	India
c	CP026940.1	100%; 100%	Plasmid	F (F2:A-:B1)	2010	*E. coli* (ST58) / (ST58, ST117, ST141)	Cow	Mastitis	France
d	CP041523.1	100%; 100%	Plasmid	F (F29:A-:B10)	2014	*E. coli* (ST69) / (ST69)	Human	Sepsis	USA
E	CP041032.1	100%; 100%	Plasmid	F (F31:A4:B1)	2018	*E. coli* (ST410) / (ST131, ST200)	Dog	Cellulitis	Portugal
	CP041338.1	100%; 100%	Plasmid	F (F36:A1:B20)	2008	*E. coli* (ST131)	Human	NA	Sweden
	LR595886.1	100%; 100%	Plasmid	F (F2:A1:B-)	2014	*E. coli* (NA)	Human	Healthy	UK
f	MK295830.1	100%; 100%	Plasmid	F (F1:A6:B20)	2007	*E. coli* (ST131) / (ST10, ST405, ST648, ST1193)	Human	NA	Israel
	MK295829.1	100%; 100%	Plasmid	F (F1:A6:B20)	2015	*E. coli* (ST131)	Dog	NA	Israel
	MK295827.1	100%; 100%	Plasmid	F (F1:A6:B20)	2015	*E. coli* (ST131)	Human	NA	Israel
G	MK295830.1	100%; 100%	Plasmid	F (F1:A6:B20)	2007	*E. coli* (ST131) /(ST38)	Human	NA	Israel
h	CP021734.1	84%; 100%	Plasmid	N (A2:N5:J-)	NA	*E. coli* (ST617) / (ST127)	NA	NA	NA
I	CP041557.1	100%; 100%	Plasmid	F (F2:A1:B-)	NA	*E. coli* (ST131) / (ST131)	Human	NA	USA

	CP024721.1	100%; 100%	Plasmid	F (F2:A1:B1)	2015	*E. coli* (ST131)	Yak	NA	China
	CP024718.1	100%; 100%	Plasmid	F (F2:A1:B1)	2015	*E. coli* (ST131)	Yak	NA	China
J	CP021289.1	100%; 100%	Plasmid	F (F51:A-:B10)	2010	*E. coli* (ST95) / (ST73, ST95, ST550)	Human	Sepsis	Australia
k*	CP023845.1	100%; 100%	Plasmid	F (F22:A1:B20)	2009	*E. coli* (ST131) / (ST131, ST405)	Human	Post UTI	Sweden
	LC056386.1	100%; 100%	Contig	-	2013	*E. coli* (NA)	Environment	NA	India
	CP028322.1	100%; 100%	Plasmid	F (F73:A-:B-)	2000	*E. coli* (ST73)	Human	NA	USA
	CP012632.1	100%; 100%	Plasmid	F (F2:A-:B-)	2008	*E. coli* (ST95)	Human	Sepsis	USA
L	CP041523.1	100%; 100%	Plasmid	F (F29:A-:B10)	2014	*E. coli* (ST69) / (ST69)	Human	Sepsis	USA
m	CP032145.1	100%; 100%	Chromosome	-	2015	*E. coli* (ST38) / (ST38)	Gull	NA	Turkey
	CP040263.1	100%; 100%	Chromosome	-	2014	*E. coli* (ST38)	Mollusk	NA	Norway
	CP040390.1	100%; 100%	Chromosome	-	2016	*E. coli* (ST38)	Gull	NA	USA
n	CP016546.1	97%; 99.97%	Chromosome	-	2015	*E. coli* (ST359) / (ST1193)	Human	Healthy	Netherlands
o	CP019014.1	100%; 100%	Plasmid	F (F2:A1:B-)	2013	*E. coli* (ST131) / (ST773)	Human	NA	USA
p	CP041393.1	100%; 100%	Plasmid	F (F36:A4:B-)	2018	*E. coli* (ST617) / (ST167)	Dog	NA	USA
	MG764548.1	100%; 100%	Plasmid	F (F33:A4:B-)	NA	*E. coli* (NA)	NA	NA	China
	CP019076.1	100%; 100%	Plasmid	F (F36:A4:B-)	2013	*E. coli* (ST617)	Human	NA	China
	CP021737.1	100%; 100%	Plasmid	F (F36:A4:B-)	NA	*E. coli* (ST617)	NA	NA	NA
	KU043115.1	100%; 100%	Plasmid	F (F36:A4:B-)	2013	*E. coli* (ST617)	Human	UTI	China
q	CP023854.1	100%; 99%	Plasmid	F (F48:A1-:B49)	2009	*E. coli* (NA) /(ST405)	Human	UTI	Sweden
r	CP014320.1	100%; 99%	Plasmid	F (F36:A-:B32)	2007	*E. coli* (ST131) /(ST448)	Human	UTI	USA

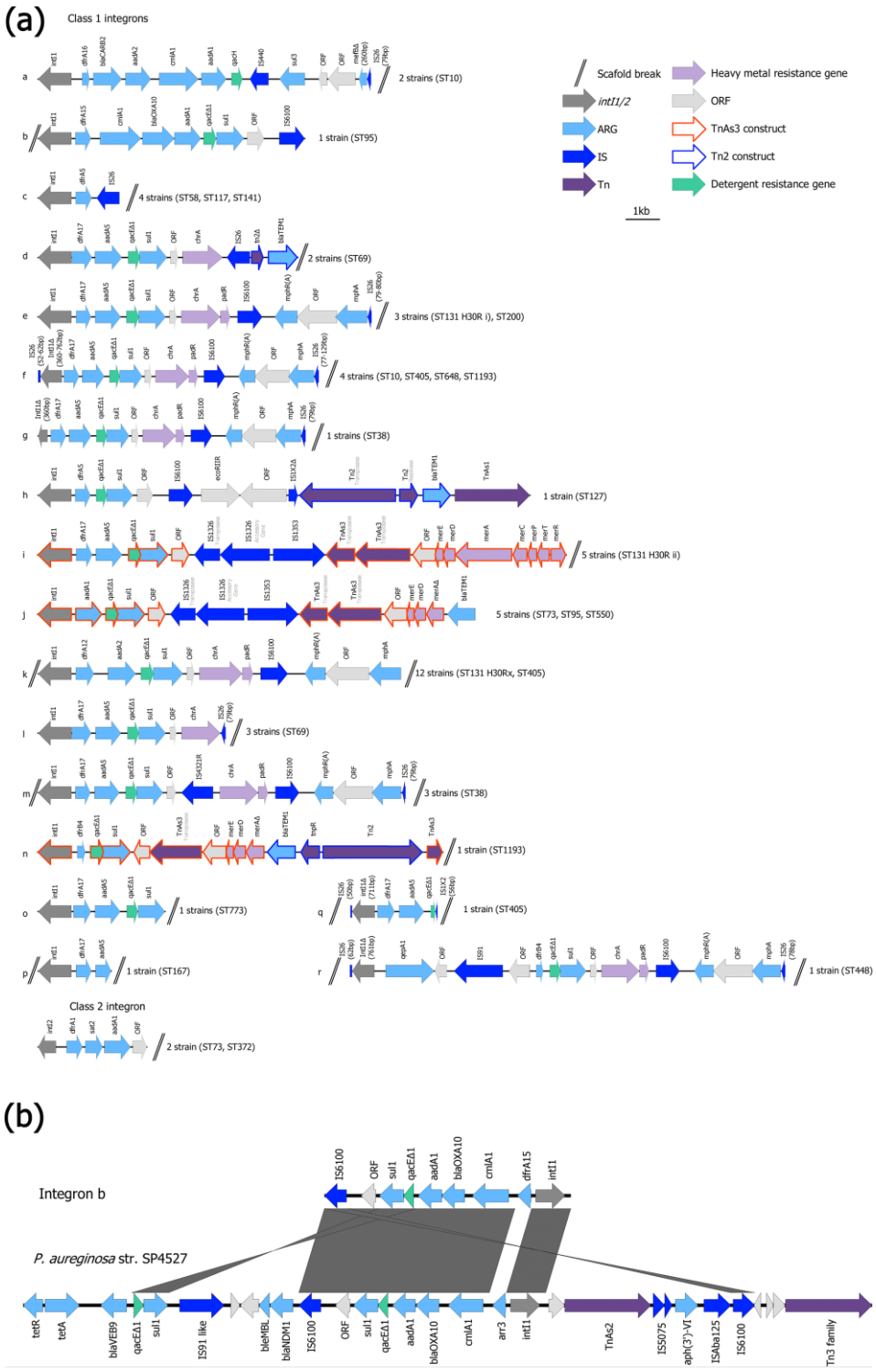


Figure 5.6: Class 1 and class 2 integron structures. a) Class 1 (a–r) and class 2 integrons structures identified in this collection. ST131 H30R i and ii refer to the two H30R subclades resolved by the SNP-based phylogeny presented in Figure 5.2(b). **b)** EasyFig comparison of integron b and the chromosome-associated integron structure from *Pseudomonas aeruginosa* strain SP4527.

5.7.4.3. VAGs

ExPEC employ a range of VAGs that enable ascension, colonization, and persistence within the urinary tract. Like ARGs, many VAGs are also carried on MGEs, including plasmids as well as chromosomally situated PAIs (15). Thus, the virulence potential of ExPEC strains is constantly evolving and tracking VAGs improves understanding of ExPEC pathogenesis. Using the VirulenceFinder database, a total of 84 VAGs were identified in this collection, with the number of VAGs per isolate spanning from 2 to 36, with a mean of 20 and a median of 19.

To combat the iron-scarcity encountered in the urinary tract, ExPEC have acquired a number of siderophore and ferrous iron uptake systems to scavenge Fe^{3+} and Fe^{2+}, which are vital for growth, persistence and establishing full virulence (15). ExPEC-associated iron-acquisition systems include yersiniabactin (*ybt*, *irp*, *fyuA*), salmochelin (*iro*), aerobactin (*iuc*, *iut*), ferric citrate transport system (*fec*) and the Sit ferrous iron utilization system (*sit*). In this study, the most common iron-acquisition genes were *fyuA*, *irp2* and *sitA* (all *n*=61, 91%), followed by *iucD* and *fecA* (both *n*=45; 67%). The high carriage of *fyuA* is consistent with previous reports tracking VAGs in ExPEC (14, 107, 108). While there are contradicting reports regarding *fyuA* and mortality (107, 109, 110), immunization with FyuA has been shown to protect against pyelonephritis in mice (111).

Several adhesins and fimbriae demonstrate specificity for binding uroepithelium. Of these, the type 1 fimbriae and P-fimbriae are most prevalent in UTI-associated ExPEC (15). Type 1 fimbriae bind uroepithelial associated α-d-mannosylated proteins via fimbrial adhesin H (*fimH*). Thus, the *fimH* gene plays a pivotal role in ExPEC urothelial adhesion (15). P-fimbriae (*pap* genes) bind α-d-galactopyranosol(1–4)-β-d-galactopyranoside-containing receptors found in the upper urinary tract and, therefore, have been associated with pyelonephritis (112). Here, the most common adhesins identified were *fimH* (*n*=61; 91%), *papB* (*n*=47; 70%) and *papI* (*n*=45; 67%). However, other P-fimbriae operon genes were less common, such as *papC*, *papD*, *papJ* and *papK* (each *n*=20; 30%), *papF* (*n*=21; 31%), *papH* and *papG* (both *n*=19; 28%), and *papA* (*n*=17; 25%).

In addition to factors facilitating adhesion and growth, ExPEC express toxins contributing to host tissue damage, as well as immune evasion molecules (113). Regarding toxins, several serine protease autotransporters of *Enterobacteriaceae* (SPATE) genes were identified including *sat* (secreted autotransporter toxin) (*n*=39; 58%), which has been shown to compromise gap junctions in uroepithelium (112), *vat* (*n*=18; 27%) and *pic* (*n*=8, 12%). Enterotoxins *senB* (*n*=21; 31%) and *astA* (*n*=4; 6%) were also observed. Regarding immune evasion, VAGs encoding

increased serum survival protein (*iss*), VirB5-like protein TraT (*traT*) and outer-membrane protein T (*ompT*) where identified in 52 (78%), 50 (75%) and 5 (7%) isolates, respectively.

Despite studies reporting various VAGs linked to pyelonephritis and sepsis (15, 114, 115), we found no single gene nor plasmid replicon strongly associated with a particular pathology (correlation heat map in Appendix 12), nor was there a significant difference between the number of VAGs present and uro-clinical syndrome (Figure 5.7a). The discrepancy with the literature may have arisen due to sampling based on trimethoprim resistance and/or the relatively small sample size, limiting statistical power. Nevertheless, an MDS analysis showed some clustering of STs based on virulence profiles (Figure 5.7b). Notably, isolates formed two distinct diagonal groups, which could be attributed to the distribution of *pap* genes with *papC*, *papD*, *papJ* and *papK* present in all isolates within the distal group and absent in all isolates in the top group. Other *pap* genes were also more prevalent in the bottom group including *papGII* (0% top, 78% bottom), *papGIII* (0, 17%), *papH* (0, 96%), *papE* (4%, 0%) and *papF* (4, 96%). Also of note was that the ST131 H30 isolates clustered with two emerging pathogen STs, ST405 (116) and ST1193 (117), as well an ST550 isolate (same CC as ST1193 – CC14) and one ST73 isolate that deviated from other ST73 isolates (Figure 5.7b, red area). The ST131 H30 isolates fell under virotype C (*afa* operon (−), *iroN* (−), *ibeA* (−), *sat* (+)), which is the most represented ST131 virotype internationally and associated with a higher frequency of infection (118). The other isolates within this cluster could also be categorized as virotype C. The only ST131 H27 isolate did not cluster with its ST and carried the molecular predictors of EAEC/ExPEC hybrid strains (*aatA*, *aggR*, *fyuA*), as did one ST200 isolate. The EAEC isolates sourced from outside this collection (Figure 5.7b, blue arrows) all deviated from ExPEC isolates of the same ST; however, they did not form a separate cluster and each carried ExPEC-associated genes including *fyuA* (Figure 5.4). These EAEC isolates were not previously described as hybrid strains and originated from cases of diarrhoeal disease in England (40). However, hybrid EAEC/ExPEC strains are known to cause UTI outbreaks (119, 120). EAEC strains carry most of their virulence cargo on plasmids and are renowned for prolific biofilm formation (121). Indeed, EAEC/ExPEC hybrid strains have shown significantly higher levels of biofilm formation, as well as enhanced adhesion to uroepithelium cells, compared to non-hybrids (122).

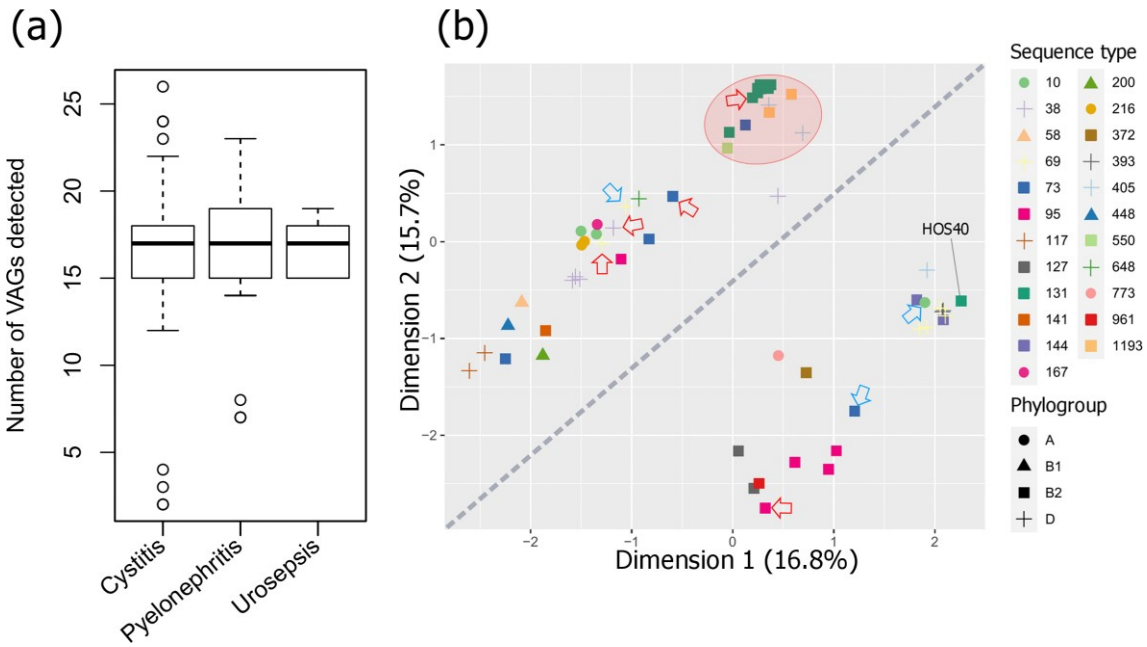


Figure 5.7: Statistical analysis of VAGs detected in the ExPEC collection. a) Box plot of uro-clinical syndrome versus the number of VAGs. Genes in operons *cvaABC*(i), *eitABCD*, *papABCDEFGHIJK*, *aafABCD*, *aggABCDR* and *mchBCF* were counted as one if at least one gene was present. **b)** MDS analysis on detected VAGs. Additional ExPEC isolates sourced from outside this collection are indicated by red arrows, additional EAEC isolates are indicated by blue arrows. The diagonal line splits isolates into a *pap* gene rich group (below the line) and a *pap* gene poor group (above the line). The red area highlights the ST131 H30 isolates cluster with two emerging pathogen STs, ST405, ST1193, as well an ST550 isolate and one ST73 isolate. The percentage of total explained variation for each dimension is indicated in parentheses after each axis label.

5.7.4.4. PAIs

The PAI marker genes *malX* and *usp* were seen in 44 (66 %) and 38 (57 %) isolates, respectively. The acquisition of PAIs via horizontal transfer can vastly contribute to the evolution of *E. coli* pathogens as they contain potent combinations of VAGs, including the aforementioned P-fimbriae, salmochelin, aerobactin and yersiniabactin operons, as well as shigella-fimbriae (S-fimbriae, *sfa*) and haemolysin toxin (*hly*) operons (123, 124). As such, to indicate the presence of ExPEC-associated PAIs, we used short-reads derived from each isolate mapped to reference sequences PAI-I_{CFT073}, PAI-II_{536} and PAI-III_{536} and PAI-IV_{536} (also known as HPI) was also screened for; however, possibly due to allelic differences, the read-depth was relatively lower for this PAI (Appendix 13) so blast was used in conjunction with assembled draft genomes to indicate its presence. While all isolates had at least partial hits to genes and gene clusters within PAI-I_{CFT073}, PAI-II_{536} and PAI-III_{536}, most demonstrated extensive deletions of regions, and few had near intact examples (Figure 5.8). The most represented VAG operon was aerobactin in PAI-I_{CFT073} and the least were S-fimbriae in PAI-III_{536}. However, PAI-IV_{536} – which consists of the yersiniabactin operon (*ybt* genes, *irp1*, *irp2*, *fyuA*; total 11 genes), a P4-like integrase and an uncharacterized

protein YeeI nestled between $tRNA_{Ser}$ and $tRNA_{Asn}$ – was identified over >95% length at >98% identity in 58 isolates (Appendix 14). In addition to being a potent siderophore system, yersiniabactin can protect against copper toxicity, redox-based phagocytosis and is a prerequisite for sepsis in some *E. coli* strains (19).

5.7.4.5. ColV plasmid markers

ColV plasmids are considered a defining feature of APEC (125); however, they have also been associated with human ExPEC strains (102, 126). While the expression of colicin V may benefit ExPEC strains by reducing competition for nutrients (127), ColV plasmids are intriguing in that they share high sequence homology to PAIs, and it is theorized that they and other plasmids and phages are the progenitors of ExPEC-associated PAIs (123). As ColV plasmids are heterogeneous in their VAGs and pMLST, Liu *et al*. (67) defined ColV plasmids as having at least one gene from four or more of the following: (i) *cvaABC* and *cvi* (ColV operon); (ii) *sitABCD* (ferrous iron utilization system); (iii) *hlyF* (regulator of outer membrane vesicle biogenesis) and *ompT*; (iv) *iucABCD* and *iutA* (aerobactin operon); (v) *iroBCDEN* (salmochelin operon); and (vi) *etsABC* (novel ABC transport system). In this study, Ariba gene screening identified seven isolates (10%) meeting this criterion – HOS66 (ST73, cystitis), HOS45 (ST73, cystitis), HOS74 (ST95, cystitis), HOS17 (ST117, pyelonephritis), HOS36 (ST117, pyelonephritis), HOS48 (ST58, cystitis) and HOS53 (ST448, cystitis). Three of these isolates (HOS17, HOS36 and HOS48) also carried the aforementioned integron c (*dfrA5*-IS26) that is used to track ColV plasmids containing complex AMR regions. Future long-read sequencing studies are needed to provide deeper insight into these initial observations (plasmid short-read mapping is provided in Appendix 15).

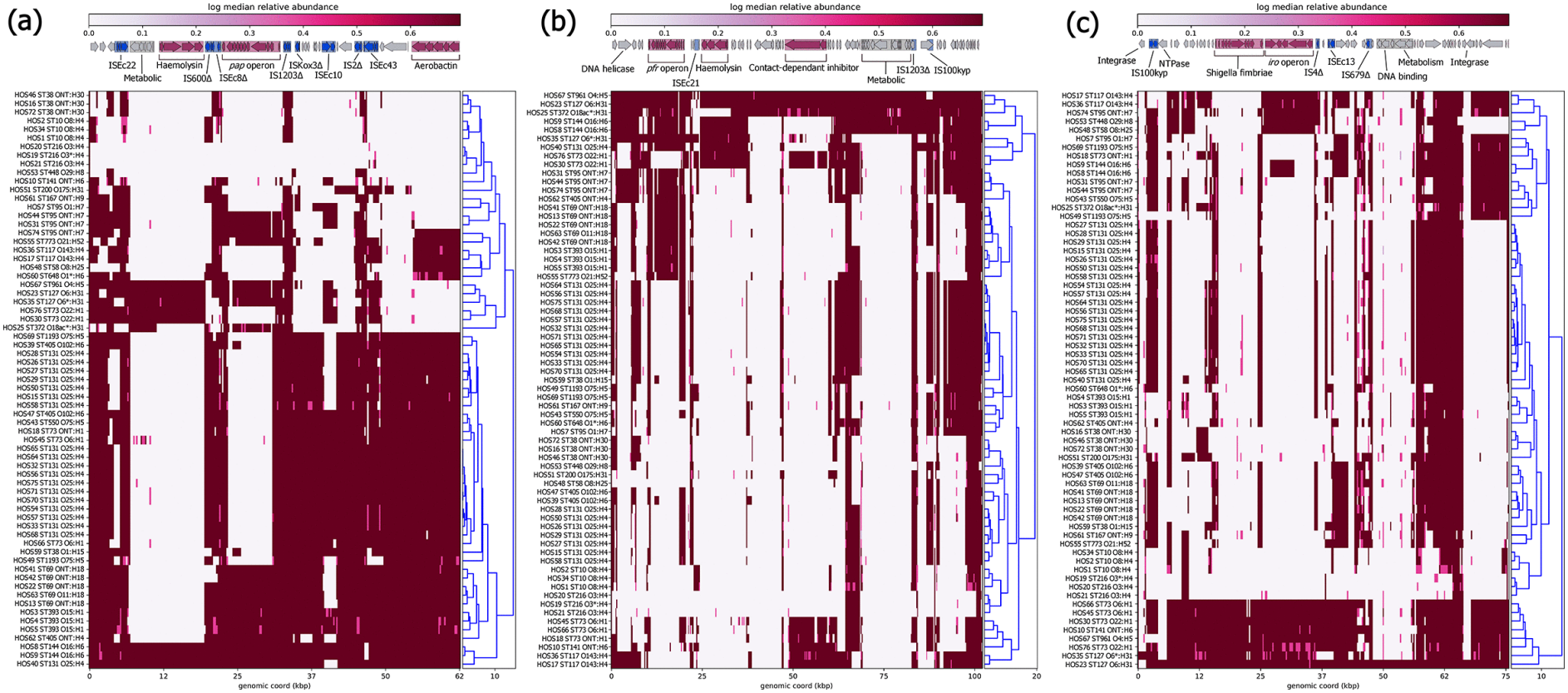


Figure 5.8: Mapping of short-reads indicating the presence of ExPEC-associated PAIs. Clustering of rows is based on the similarity between isolate PAI coverage profiles. **a)** Schematic and heatmap for PAI I_{CFT072} coverage. **b)** Schematic and heatmap for PAI II_{536} coverage. **c)** Schematic and heatmap for PAI III_{536}coverage. Trees alongside the heatmaps were reconstructed by hierarchical clustering using Euclidean agglomeration method.

5.7.5. SNP analysis of serial patient isolates

Phenotypic diversification via adaptive evolution is driven by HGT and by mutation (128). A single SNP-mediated point mutation can lead to an amino acid replacement with significant adaptive effects, as witnessed in single amino acid replacements in *parC/E* and *gyrA* genes causing fluoroquinolone resistance (24, 25), and in *fimH*, which increases binding to uroepithelium and biofilm formation (129). Such adaptive mutations can increase bacterial fitness and provide rapid adaption to niche-specific conditions (128). ExPEC are exposed to a variety of niches in addition to their primary reservoir, the intestinal tract, and as a result may be under greater selective pressure for adaptive mutations. Indeed, UTI-associated ExPEC are known to have significantly higher mutation rates compared to other *E. coli* pathotypes (130). While studies have been conducted comparing SNP-mediated point mutations in UTI and faecal isolates (131), relatively little is known about the SNPs that occur in recurrent UTIs, particularly when subsequently more serious sequelae occur. In this study, 13 patients had more than one presentation of symptomatic UTI during the study period (indicated by letters in Figure 5.2). Six patients had subsequent UTIs caused by the same ST, and seven patients had different STs. Here, we focus on the patients who had multiple infections with the same ST (Figure 5.9), and on SNP-mediated mutations with a higher probability of inducing functional changes as predicted by the BLOSUM62 substitution matrix (132). A catalogue of all SNPs can be viewed in Appendix 16. The BLOSUM62 matrix provides a scoring system for aligning amino acid residues based on their observed substitutions in a large set of protein alignments (133). It assigns a numerical score to pairs of amino acids, indicating the likelihood of their substitution during evolution. It was chosen due to its widespread acceptance and extensive validation. It provides a well-balanced scoring system that accommodates both closely and distantly related sequences. High availability of this matrix in widely used bioinformatics tools and databases ensures compatibility and comparability across analyses.

Patient A was a 37-year-old female who in the first instance presented with cystitis with an ST10 isolate (HOS1) as the causative agent and then pyelonephritis (HOS2; ST10) 63 days later. The ST10 isolates had a difference of 29 SNPs, with most occurring in genes involved in DNA processing, transport, and virulence (Figure 5.9). However, mutations with negative BLOSUM62 scores (indicating a higher probability that the original function is not conserved) occurred most frequently in VAGs and DNA processing. Potentially increasing the significance of the observed mutations, another patient, a 69-year-old male living in the same postcode, presented with pyelonephritis 16 days after patient A's second presentation. The causative agent was also an ST10 (HOS34), which shared 71% (n = 24) of the same SNPs found in HOS2. Both HOS2 and

HOS34 had the same SNPs causing non-synonymous amino acid substitutions with negative BLOSUM scores in the following genes: *irp1* (D2717A), *ybtS* (P184A), *astA* (G103V), *aroC* (G152E), *rnhB* (R98L), *hrpB* (G644E), *traC* (G644W), *lsrA* (E296G) and *recC* (S794C). Interestingly, *recC* was previously shown to be under positive selective pressure for adaptive mutations (130). Though not present in HOS34, HOS2 possessed a SNP causing A432D (negative BLOSUM62 score) in DNA-repair protein Mfd (mutation-frequency-decline). Mfd promotes mutagenesis and accelerates AMR development to multiple antibiotic classes (134). Fascinatingly, point mutations produce Mfd variants that express greater activity than wild-type (135).

Patient B was a 21-year-old-female who first presented with cystitis (HOS3) then pyelonephritis 284 days later (HOS4, 13 SNP difference), then pyelonephritis again 128 days later (HOS5, 19 SNP difference from HOS3). The causative agent in each episode was ST393; however, unlike the high percentage of conserved SNPs in ST10 pyelonephritis isolates, HOS4 and HOS5 ST393 only shared 23 % of all SNPs against the HOS3 reference, and only shared two negative BLOSUM62 score amino acid substitutions in chromosome partition protein MukF (F401V and T394H). Most SNPs occurred in intergenic regions, which may be significant as these regions often contain riboswitches, promoters, terminators and regulator binding sites, and SNPs can cause significant phenotypic changes (136).

Patient C was a 64-year-old female with cystitis and two samples were taken on the same day. Both isolates (HOS8 and HOS9) were ST144 and differed by five SNPs. While no SNPs occurred in VAGs (Figure 5.9), two mutations with potential for functional change were identified in transposase TnpB (G11R) and environmental stress protein Ves (G7R). Conversely, in patient H, a 67-year-old male who presented with cystitis (HOS33; ST131 H30Rx) and then cystitis again a week later (HOS34; ST131 H30Rx), the two causative agents were indistinguishable.

Patient F was a 72-year-old female with cystitis followed by urosepsis 181 days later (HOS26; ST131 H30R; 53 SNPs or 20 SNPs without recombination), followed by urosepsis 101 days later (HOS27; ST131 H30R; 61 or 27 SNPs) and then a subsequent sample was taken 3 days later (HOS28; ST131 H30R; 62 or 26 SNPs). The vast majority of SNPs were a result of recombination, which is 50–100 times more likely to occur than mutations (137) and occurred in aerobactin biosynthesis protein *iucC* and iron transport system genes *sitABC*. However, three-point mutations with potential change to function were identified in all urosepsis isolates: R18L in capsule protein KpsM, R268L in molybdoenzyme biosynthesis protein MoeA and M58R in molybdoenzyme chaperon protein YcdY. Intriguingly, HOS50 (also ST131 H30R), originating from a 50-year-old patient presenting with pyelonephritis 21 days prior to patient F's initial

presentation, also possessed these three precise amino acid substitutions produced by the same SNPs. These identical substitutions may be indicative of a shared source, particularly as both patients resided in the same postcode, or may be examples of pathoadaptive mutations that increase fitness in the upper urinary tract. However, no recombination events were detected in HOS50.

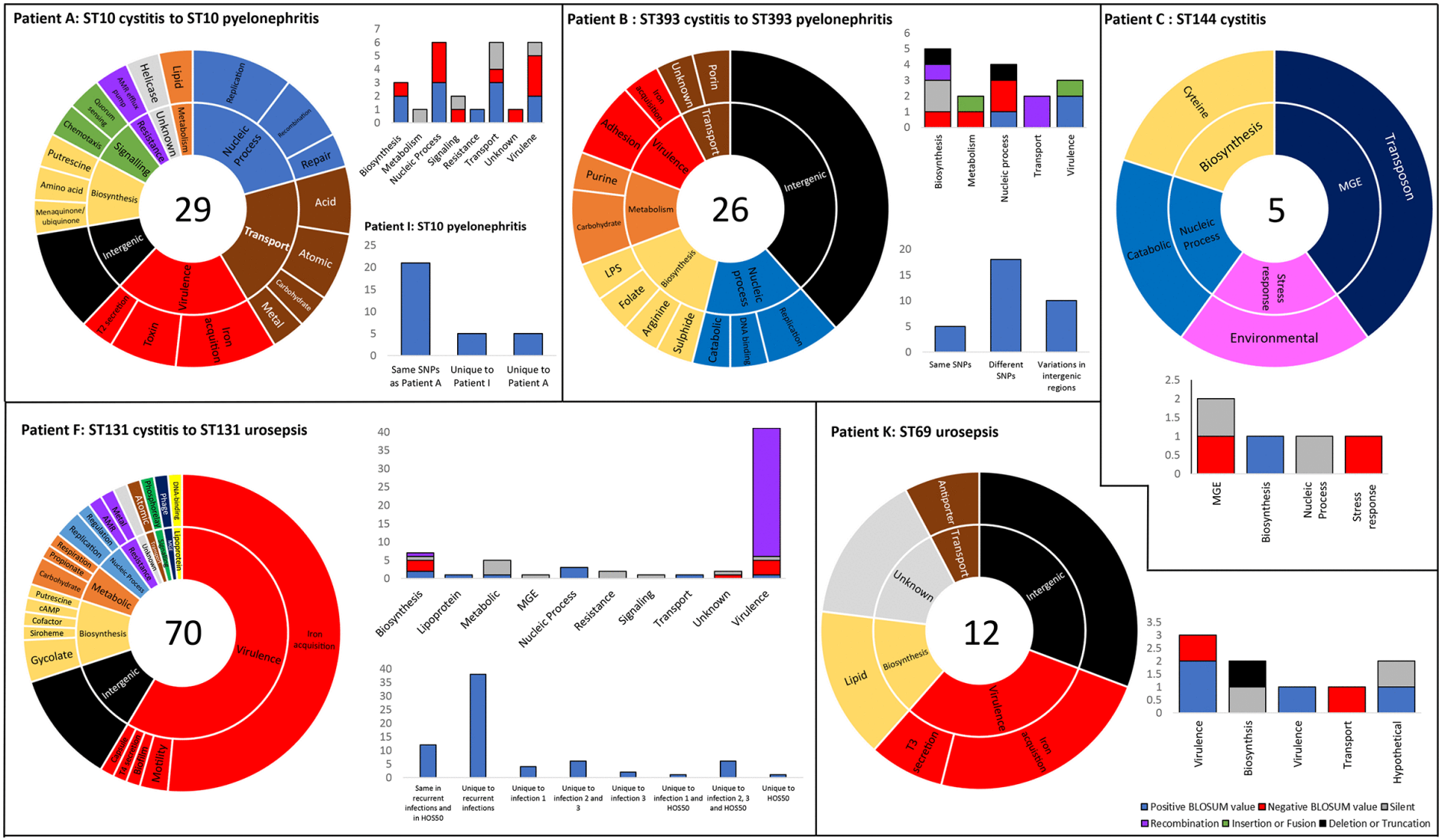

Patient A: ST10 cystitis to ST10 pyelonephritis
29
Nucleic Process
Replication
Recombination
Repair
Transport
Acid
Atomic
Carbohydrate
Metal
Iron acquition
Virulence
T2 secretion
Toxin
Intergenic
Biosynthesis
Amino acid
Menaquinone/ ubiquinone
Putrescine
Signalling
Chemotaxis
Quorum sensing
AMR efflux pump
Resistance
Helicase
Unknown
Metabolism
Lipid
Biosynthesis
Metabolism
Nucleic Process
Signaling
Resistance
Transport
Unknown
Virulene
Patient I: ST10 pyelonephritis
Same SNPs as Patient A
Unique to Patient I
Unique to Patient A
Patient B : ST393 cystitis to ST393 pyelonephritis
26
Intergenic
Transport
Porin
Unknown
Iron acquisition
Virulence
Adhesion
Metabolism
Purine
Carbohydrate
LPS
Folate
Arginine
Sulphide
Biosynthesis
Nucleic process
DNA binding
Replication
Catabolic
Biosynthesis
Metabolism
Nucleic process
Transport
Virulence
Same SNPs
Different SNPs
Variations in intergenic regions
Patient C : ST144 cystitis
5
Transposon
MGE
Stress response
Environmental
Nucleic Process
Catabolic
Biosynthesis
Cyteine
MGE
Biosynthesis
Nucleic Process
Stress response
Patient F: ST131 cystitis to ST131 urosepsis
70
Iron acquisition
Virulence
Motility
Biofilm
T2 secretion
Capsule
Intergenic
Gycolate
Biosynthesis
Cofactor
Siroheme
cAMP
Putrescine
Carbohydrate
Propionate
Respiration
Replication
Regulation
AMR
Metal
Metabolic
Nucleic Process
Resistance
Unknown
Signaling
Lipoprotein
Phosphonate
Atomic
Phage
DNA binding
Biosynthesis
Lipoprotein
Metabolic
MGE
Nucleic Process
Resistance
Signaling
Transport
Unknown
Virulence
Same in recurrent infections and in HOS50
Unique to recurrent infections
Unique to infection 1
Unique to infection 2 and 3
Unique to infection 3
Unique to infection 1 and HOS50
Unique to infection 2, 3 and HOS50
Unique to HOS50
Patient K: ST69 urosepsis
12
Intergenic
Transport
Antiporter
Unknown
Virulence
Iron acquistion
T3 secretion
Biosynthesis
Lipid
Virulence
Biosynthsis
Virulence
Transport
Hypothetical
Positive BLOSUM value
Negative BLOSUM value
Silent
Recombination
Insertion or Fusion
Deletion or Truncation

Figure 5.9: Serial patient isolates – SNP distribution and consequence to amino acid sequence. Each panel represents a UTI patient from whom multiple isolates of the same ST were derived. Sunburst graphs represent the distribution of SNPs within functional groups. The number in the centre represents the total number of unique SNPs (accumulative if more than one recurrent infection occurred). Stacked column graphs represent the consequence of SNPs to protein sequence. Colour legend for the stacked columns located bottom right corner. Unstacked column graphs represent the distribution of specific SNPs in recurrent infections from the same patient, and in the case of patient A and F, two different patients.

Lastly, patient K was a 2-year-old female who presented with urosepsis (HOS41; ST69) and subsequent urosepsis 100 days later (HOS42; ST69; 13 SNPs). Similar to isolates derived from patient B, most SNPs occurred within intergenic regions (Figure 5.9); however, R133L in yersiniabactin protein YbtT and R337S in K+/H+ antiporter CvrA may have incurred functional changes. Future characterization of all these SNP-mediated mutations is required to ascertain any functional changes.

5.8. Limitations of study

There are several limitations in this study that should be considered when interpreting the results. Firstly, the sample size was small, consisting of only 67 *E. coli* draft genomes. Consequently, this limited sample size may not accurately represent the larger population of trimethoprim resistant UTI-associated ExPEC within Australian hospitals. Secondly, the selection of isolates was based on phenotypic resistance may introduce bias and restrict the generalisability of the findings to a broader population. Additionally, the study focused on a regional hospital in Australia so the findings may not be applicable to other regions or hospitals within the country. Furthermore, the study encountered difficulties in assembling complete class 1 and class 2 integron structures due to the presence of repetitive sequences and the utilization of short-read sequencing data. This limitation hinders a comprehensive understanding of the genetic composition and arrangement of integrons, which play a crucial role in antibiotic resistance. Finally, the study relied on databases such as ResFinder and VirulenceFinder for the analysis of ARG and VAG carriage which have their own limitations, including incomplete or outdated entries, that can impact the accuracy and completeness of the analysis.

In conclusion, this study highlights the concern of antibiotic resistance genes in ExPEC isolates. The presence of class 1 and 2 integrons, along with their association with ARG, further emphasizes the potential for multidrug resistance dissemination. Additionally, the identification of numerous VAGs underscores the evolving nature of ExPEC strains and their ability to cause UTIs. These findings emphasize the urgent need for comprehensive surveillance, prudent use of

antimicrobials, and effective infection control measures to mitigate the emergence and spread of multidrug-resistant ExPEC strains, thereby safeguarding public health.

5.9. References

1. Simmering JE, Tang F, Cavanaugh JE, Polgreen LA, Polgreen PM. The increase in hospitalizations for urinary tract infections and the associated costs in the United States, 1998–2011. Open Forum Infect Dis. 2017;4:ofw281. doi: 10.1093/ofid/ofw281.
2. Medical Technology Group Admissions of Failure: the Truth About Unplanned NHS Admissions in England (https://mtg.org.uk/wp-content/uploads/2016/07/Admissions-of-Failure-2015.pdf) London: Medical Technology Group; 2015.
3. Australian Commission on Safety and Quality in Health Care Atlas 2017 - 1. Chronic Disease and Infection: Potentially Preventable Hospitalisations (www.safetyandquality.gov.au/our-work/healthcare-variation/atlas-2017/atlas-2017-1-chronic-disease-and-infection-potentially-preventable-hospitalisations) Sydney: Australian Commission on Safety and Quality in Health Care; 2017.
4. Australian Commission on Safety and Quality in Health Care AURA 2019: Third Australian Report on Antimicrobial Use and Resistance in Human Health (www.safetyandquality.gov.au/publications-and-resources/resource-library/aura-2019-third-australian-report-antimicrobial-use-and-resistance-human-health) Sydney: Australian Commission on Safety and Quality in Health Care; 2019.
5. Abbo LM, Hooton TM. Antimicrobial stewardship and urinary tract infections. Antibiotics. 2014;3:174–192. doi: 10.3390/antibiotics3020174.
6. Foxman B, Brown P. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect Dis Clin North Am. 2003;17:227–241.
7. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, et al. Global extraintestinal pathogenic Escherichia coli (ExPEC) lineages. Clin Microbiol Rev. 2019;32:e00135-18. doi: 10.1128/CMR.00135-18.
8. Yamaji R, Rubin J, Thys E, Friedman CR, Riley LW. Persistent pandemic lineages of uropathogenic Escherichia coli in a college community from 1999 to 2017. J Clin Microbiol. 2018;56:e01834-17. doi: 10.1128/JCM.01834-17.
9. Manges AR, Johnson JR. Reservoirs of extraintestinal pathogenic Escherichia coli . Microbiol Spectr. 2015;3:UTI-0006-2012. doi: 10.1128/microbiolspec.UTI-0006-2012.
10. Kantele A, Lääveri T, Mero S, Häkkinen IMK, Kirveskari J, 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. Clin Infect Dis. 2020;70:210–218. doi: 10.1093/cid/ciz182.
11. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. Microbes Infect. 2003;5:449–456. doi: 10.1016/S1286-4579(03)00049-2.
12. Epstein L, Dantes R, Magill S, Fiore A. Varying estimates of sepsis mortality using death certificates and administrative codes — United States, 1999–2014. MMWR Morb Mortal Wkly Rep. 2016;65:342–345. doi: 10.15585/mmwr.mm6513a2.

13. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. JAMA. 2017;318:1241–1249. doi: 10.1001/jama.2017.13836.
14. Daga AP, Koga VL, Soncini JGM, de Matos CM, Perugini MRE, et al. Escherichia coli bloodstream infections in patients at a university hospital: virulence factors and clinical characteristics. Front Cell Infect Microbiol. 2019;9:191. doi: 10.3389/fcimb.2019.00191.
15. Dale AP, Woodford N. Extra-intestinal pathogenic Escherichia coli (ExPEC): disease, carriage and clones. J Infect. 2015;71:615–626. doi: 10.1016/j.jinf.2015.09.009.
16. Micenková L, Bosák J, Vrba M, Ševčíková A, Šmajs D. Human extraintestinal pathogenic Escherichia coli strains differ in prevalence of virulence factors, phylogroups, and bacteriocin determinants. BMC Microbiol. 2016;16:218. doi: 10.1186/s12866-016-0835-z.
17. Johnson TJ, Nolan LK. Pathogenomics of the virulence plasmids of Escherichia coli . MMBR. 2009;73:750–774. doi: 10.1128/MMBR.00015-09.
18. Peirano G, Mulvey GL, Armstrong GD, Pitout JDD. Virulence potential and adherence properties of Escherichia coli that produce CTX-M and NDM β-lactamases. J Med Microbiol. 2013;62:525–530. doi: 10.1099/jmm.0.048983-0.
19. Johnson JR, Magistro G, Clabots C, Porter S, Manges A, et al. Contribution of yersiniabactin to the virulence of an Escherichia coli sequence type 69 (“clonal group A”) cystitis isolate in murine models of urinary tract infection and sepsis. Microb Pathog. 2018;120:128–131. doi: 10.1016/j.micpath.2018.04.048.
20. Hancock V, Ferrières L, Klemm P. The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious Escherichia coli in human urine. Microbiology. 2008;154:167–175. doi: 10.1099/mic.0.2007/011981-0.
21. Magistro G, Hoffmann C, Schubert S. The salmochelin receptor IroN itself, but not salmochelin-mediated iron uptake promotes biofilm formation in extraintestinal pathogenic Escherichia coli (ExPEC) Int J Med Microbiol. 2015;305:435–445. doi: 10.1016/j.ijmm.2015.03.008.
22. Alam M, Bastakoti B. Therapeutic guidelines: antibiotic, version 15. Aust Prescr. 2015;38:137. doi: 10.18773/austprescr.2015.049.
23. Fasugba O, Mitchell BG, Mnatzaganian G, Das A, Collignon P, et al. Five-year antimicrobial resistance patterns of urinary Escherichia coli at an Australian tertiary hospital: time series analyses of prevalence data. PLoS One. 2016;11:e0164306. doi: 10.1371/journal.pone.0164306.
24. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev. 2014;27:543–574. doi: 10.1128/CMR.00125-13.
25. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, et al. The epidemic of extended-spectrum-β-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. mBio. 2013;4:e00377-13. doi: 10.1128/mBio.00377-13.
26. Paltansing S, Kraakman MEM, Ras JMC, Wessels E, Bernards AT. Characterization of fluoroquinolone and cephalosporin resistance mechanisms in Enterobacteriaceae isolated in a Dutch teaching hospital reveals the presence of

an Escherichia coli ST131 clone with a specific mutation in parE. J Antimicrob Chemother. 2013;68:40–45. doi: 10.1093/jac/dks365.
27. Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, et al. Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli . J Infect Dis. 2013;207:919–928. doi: 10.1093/infdis/jis933.
28. 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;213:6–13. doi: 10.1093/infdis/jiv429.
29. Zilberberg MD, Shorr AF. Secular trends in gram-negative resistance among urinary tract infection hospitalizations in the United States, 2000–2009. Infect Control Hosp Epidemiol. 2013;34:940–946. doi: 10.1086/671740.
30. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. J Antimicrob Chemother. 2017;72:2145–2155. doi: 10.1093/jac/dkx146.
31. Bajaj P, Singh NS, Virdi JS. Escherichia coli β-lactamases: what really matters. Front Microbiol. 2016;7:417. doi: 10.3389/fmicb.2016.00417.
32. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. Clin Microbiol Rev. 2018;31:e00088-17. doi: 10.1128/CMR.00088-17.
33. Clinical and Laboratory Standards Institute Performance Standards for Antimicrobial Susceptibility Testing. Wayne, PA: Clinical and Laboratory Standards Institute; 2017.
34. Clinical and Laboratory Standards Institute Performance Standards for Antimicrobial Disk Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard, M02-A10. 10th edn. Wayne, PA: Clinical and Laboratory Standards Institute; 2009.
35. Reid CJ, Wyrsch ER, Chowdhury PR, Zingali T, Liu M, et al. Porcine commensal Escherichia coli: a reservoir for class 1 integrons associated with IS26. Microb Genom. 2017;3:mgen.0.000143. doi: 10.1099/mgen.0.000143.
36. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–1534. doi: 10.1093/molbev/msaa015.
37. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15. doi: 10.1093/nar/gku1196.
38. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, et al. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res. 2017;27:1437–1449. doi: 10.1101/gr.216606.116.
39. Nascimento V, Day MR, Doumith M, Hopkins KL, Woodford N, et al. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of enteroaggregative Escherichia coli isolated from cases of diarrhoeal disease in England, 2015–16. J Antimicrob Chemother. 2017;72:3288–3297. doi: 10.1093/jac/dkx301.
40. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256–W259. doi: 10.1093/nar/gkz239.

41. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3693. doi: 10.1093/bioinformatics/btv421.
42. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, et al. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics. 2018;34:292–293. doi: 10.1093/bioinformatics/btx610.
43 Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–2069. doi: 10.1093/bioinformatics/btu153.
44. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol. 2012;50:1355–1361. doi: 10.1128/JCM.06094-11.
45. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol. 2015;53:2410–2426. doi: 10.1128/JCM.00008-15.
46. 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;5:58–65. doi: 10.1111/1758-2229.12019.
47. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. blast+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi: 10.1186/1471-2105-10-421.
48. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3:mgen.0.000131. doi: 10.1099/mgen.0.000131.
49. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–2644. doi: 10.1093/jac/dks261.
50. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, et al. In silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58:3895–3903. doi: 10.1128/AAC.02412-14.
51. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli . J Clin Microbiol. 2014;52:1501–1510. doi: 10.1128/JCM.03617-13.
52. Johnson JR, Stell AL, Scheutz F, O'Bryan TT, Russo TA, et al. Analysis of the F antigen-specific papAAlleles of extraintestinal pathogenic Escherichia coli using a novel multiplex PCR-based assay. Infect Immun. 2000;68:1587–1599. doi: 10.1128/IAI.68.3.1587-1599.2000.
53. Yu G, Smith DK, Zhu H, Guan Y, Lam Tommy Tsan-Yuk. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. doi: 10.1111/2041-210X.12628.
54. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–1760. doi: 10.1093/bioinformatics/btp324.
55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–2079. doi: 10.1093/bioinformatics/btp352.

56. Leray M, Knowlton N, Ho S-L, Nguyen BN, Machida RJ. GenBank is a reliable resource for 21st century biodiversity research. Proc Natl Acad Sci USA. 2019;116:22651–22656. doi: 10.1073/pnas.1911714116.
57. Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Agama Study Group. et al. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny and Escherichia core genomic diversity. Genome Res. 2019;119:251678
58. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27:1009–1010. doi: 10.1093/bioinformatics/btr039.
59. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, et al. Using genomics to track global antimicrobial resistance. Front Public Health. 2019;7:242. doi: 10.3389/fpubh.2019.00242.
60. Al-Badr A, Al-Shaikh G. Recurrent urinary tract infections management in women. Sultan Qaboos Univ Med J. 2013;13:359–367.
61. Schmiemann G, Kniehl E, Gebhardt K, Matejczyk MM, Hummers-Pradier E. The diagnosis of urinary tract infection: a systematic review. Dtsch Arzteblatt Int. 2010;107:361–367.
62. Chu CM, Lowder JL. Diagnosis and treatment of urinary tract infections across age groups. Am J Obstet Gynecol. 2018;219:40–51. doi: 10.1016/j.ajog.2017.12.231.
63. Johnson JR, Menard M, Johnston B, Kuskowski MA, Nichol K, et al. Epidemic clonal groups of Escherichia coli as a cause of antimicrobial-resistant urinary tract infections in Canada, 2002 to 2004. Antimicrob Agents Chemother. 2009;53:2733–2739. doi: 10.1128/AAC.00297-09.
64. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, et al. Escherichia coli sequence type 131 as a prominent cause of antibiotic resistance among urinary Escherichia coli isolates from reproductive-age women. J Clin Microbiol. 2013;51:3270–3276. doi: 10.1128/JCM.01315-13.
65. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, et al. Genotypic and phenotypic characterization of Escherichia coli isolates from children with urinary tract infection and from healthy carriers. Pediatr Infect Dis J. 2013;32:543–548. doi: 10.1097/INF.0b013e31828ba3f1.
66. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, et al. Distribution of phylogenetic groups, sequence type ST131, and virulence-associated traits among Escherichia coli isolates from men with pyelonephritis or cystitis and healthy controls. Clin Microbiol Infect. 2013;19:E173–E180. doi: 10.1111/1469-0691.12123.
67. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, et al. Escherichia coli ST131-H22 as a foodborne thogen. mBio. 2018;9:mBio.00470-18. doi: 10.1128/mBio.00470-18.
68. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae . Clin Microbiol Rev. 2015;28:565–591. doi: 10.1128/CMR.00116-14.
69. Ewers C, Göttig S, Bülte M, Fiedler S, Tietgen M, et al. Genome sequence of avian Escherichia coli strain IHIT25637, an extraintestinal pathogenic E. coli strain of ST131 encoding colistin resistance determinant MCR-1. Genome Announc. 2016;4:e00863-16. doi: 10.1128/genomeA.00863-16.

70. Johnson TJ, Hargreaves M, Shaw K, Snippes P, Lynfield R, et al. Complete genome sequence of a carbapenem-resistant extraintestinal pathogenic Escherichia coli strain belonging to the sequence type 131 H30R subclade. Genome Announc. 2015;3:e00272-15. doi: 10.1128/genomeA.00272-15.
71. World Health Organization Global Action Plan on Antimicrobial Resistance (www.who.int/antimicrobial-resistance/publications/global-action-plan/en/) Geneva: World Health Organization; 2015.
72. Pearson J, Turnidge J, Franklin C, Bell J, Australian Group on Antimicrobial Resistance Prevalence of antimicrobial resistances in common pathogenic Enterobacteriaceae in Australia, 2004: report from the Australian Group on Antimicrobial Resistance. Commun Dis Intell Q Rep. 2007;31:106–112.
73. Bell JM, Gottlieb T, Daley DA, Coombs GW. Australian Group on Antimicrobial Resistance (AGAR) Australian gram-negative sepsis outcome programme (GNSOP) annual report 2017. Commun Dis Intell. 2019;43:cdi.2019.43.37. doi: 10.33321/cdi.2019.43.37.
74. Kot B. Antibiotic resistance among uropathogenic Escherichia coli . Pol J Microbiol. 2019;68:403–415. doi: 10.33073/pjm-2019-048.
75. Thänert R, Reske KA, Hink T, Wallace MA, Wang B, et al. Comparative genomics of antibiotic-resistant uropathogens implicates three routes for recurrence of urinary tract infections. mBio. 2019;10:e01977-19. doi: 10.1128/mBio.01977-19.
76. Mortazavi-Tabatabaei SAR, Ghaderkhani J, Nazari A, Sayehmiri K, Sayehmiri F, et al. Pattern of antibacterial resistance in urinary tract infections: a systematic review and meta-analysis. Int J Prev Med. 2019;10:169.
77. Critchley IA, Cotroneo N, Pucci MJ, Mendes R. The burden of antimicrobial resistance among urinary tract isolates of Escherichia coli in the United States in 2017. PLoS One. 2019;14:e0220265. doi: 10.1371/journal.pone.0220265.
78. Mulder M, Verbon A, Lous J, Goessens W, Stricker BH. Use of other antimicrobial drugs is associated with trimethoprim resistance in patients with urinary tract infections caused by E. coli . Eur J Clin Microbiol Infect Dis. 2019;38:2283–2290. doi: 10.1007/s10096-019-03672-2.
79. Seputiene V, Povilonis J, Ruzauskas M, Pavilonis A, Suziedeliene E. Prevalence of trimethoprim resistance genes in Escherichia coli isolates of human and animal origin in Lithuania. J Med Microbiol. 2010;59:315–322. doi: 10.1099/jmm.0.015008-0.
80. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48:D517–D525. doi: 10.1093/nar/gkz935.
81. Djordjevic SP, Stokes HW, Chowdhury PR. Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. Front Microbiol. 2013;4:86. doi: 10.3389/fmicb.2013.00086.
82. Cummins ML, Reid CJ, Chowdhury PR, Bushell RN, Esbert N, et al. Whole genome sequence analysis of Australian avian pathogenic Escherichia coli that carry the class 1 integrase gene. Microb Genomics. 2019;5:e000250. doi: 10.1099/mgen.0.000250.
83. Jarocki VM, Reid CJ, Chapman TA, Djordjevic SP. Escherichia coli ST302: genomic analysis of virulence potential and antimicrobial resistance mediated by mobile genetic elements. Front Microbiol. 2020;10:3098. doi: 10.3389/fmicb.2019.03098.

84. Reid CJ, Blau K, Jechalke S, Smalla K, Djordjevic SP. Whole genome sequencing of Escherichia coli from store-bought produce. Front Microbiol. 2020;10:03050. doi: 10.3389/fmicb.2019.03050.
85. Poirel L, Kämpfer P, Nordmann P. Chromosome-encoded ambler class A β-lactamase of Kluyvera georgiana, a probable progenitor of a subgroup of CTX-M extended-spectrum β-lactamases. Antimicrob Agents Chemother. 2002;46:4038–4040. doi: 10.1128/AAC.46.12.4038-4040.2002.
86. Cormier A, Zhang PLC, Chalmers G, Weese JS, Deckert A, et al. Diversity of CTX-M-positive Escherichia coli recovered from animals in Canada. Vet Microbiol. 2019;231:71–75. doi: 10.1016/j.vetmic.2019.02.031.
87. El-Badawy MF, Tawakol WM, Maghrabi IA, Mansy MS, Shohayeb MM, et al. Iodometric and molecular detection of ESBL production among clinical isolates of E. coli fingerprinted by ERIC-PCR: the first Egyptian report declares the emergence of E. coli O25b-ST131clone harboring bla GES . Microb Drug Resist. 2017;23:703–717. doi: 10.1089/mdr.2016.0181.
88. Yuan L, Liu J-H, Hu G-Z, Pan Y-S, Liu Z-M, et al. Molecular characterization of extended-spectrum β-lactamase-producing Escherichia coli isolates from chickens in Henan Province, China. J Med Microbiol. 2009;58:1449–1453. doi: 10.1099/jmm.0.012229-0.
89. Farajzadeh Sheikh A, Goodarzi H, Yadyad MJ, Aslani S, Amin M, et al. Virulence-associated genes and drug susceptibility patterns of uropathogenic Escherichia coli isolated from patients with urinary tract infection. Infect Drug Resist. 2019;12:2039–2047. doi: 10.2147/IDR.S199764.
90. Ramírez-Castillo FY, Moreno-Flores AC, Avelar-González FJ, Márquez-Díaz F, Harel J, et al. An evaluation of multidrug-resistant Escherichia coli isolates in urinary tract infections from Aguascalientes, Mexico: cross-sectional study. Ann Clin Microbiol Antimicrob. 2018;17:34. doi: 10.1186/s12941-018-0286-5.
91. Ochoa SA, Cruz-Córdova A, Luna-Pineda VM, Reyes-Grajeda JP, Cázares-Domínguez V, et al. Multidrug- and extensively drug-resistant uropathogenic Escherichia coli clinical strains: phylogenetic groups widely associated with integrons maintain high genetic diversity. Front Microbiol. 2016;7:02042. doi: 10.3389/fmicb.2016.02042.
92. Li L-G, Xia Y, Zhang T. Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. ISME J. 2017;11:651–662. doi: 10.1038/ismej.2016.155.
93. Yang S, Deng W, Liu S, Yu X, Mustafa GR, et al. Presence of heavy metal resistance genes in Escherichia coli and Salmonella isolates and analysis of resistance gene structure in E. coli E308. J Glob Antimicrob Resist. 2020;21:420–426. doi: 10.1016/j.jgar.2020.01.009.
94. Nguyen CC, Hugie CN, Kile ML, Navab-Daneshmand T. Association between heavy metals and antibiotic-resistant human pathogens in environmental reservoirs: a review. Front Environ Sci Eng. 2019;13:46. doi: 10.1007/s11783-019-1129-0.
95. Department of Primary Industries Central West Region Pilot Area Agricultural Profile (www.dpi.nsw.gov.au/__data/assets/pdf_file/0010/457588/Agricultural-profile-central-west-region.pdf) Sydney: NSW Government; 2012.
96. Knapp CW, McCluskey SM, Singh BK, Campbell CD, Hudson G, et al. Antibiotic resistance gene abundances correlate with metal and geochemical conditions in

archived Scottish soils. PLoS One. 2011;6:e27300. doi: 10.1371/journal.pone.0027300.
97. Gillings MR. Integrons: past, present, and future. Microbiol Mol Biol Rev. 2014;78:257–277. doi: 10.1128/MMBR.00056-13.
98. Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP, et al. Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in Escherichia coli of human and animal origin. PLoS One. 2010;5:e12754. doi: 10.1371/journal.pone.0012754.
99. Brolund A, Rajer F, Giske CG, Melefors Ö, Titelman E, et al. Dynamics of resistance plasmids in extended-spectrum-β-lactamase-producing Enterobacteriaceae during postinfection colonization. Antimicrob Agents Chemother. 2019;63:e02201-18. doi: 10.1128/AAC.02201-18.
100. Toulouse JL, Edens TJ, Alejaldre L, Manges AR, Pelletier JN. Integron-associated DfrB4, a previously uncharacterized member of the trimethoprim-resistant dihydrofolate reductase B family, is a clinically identified emergent source of antibiotic resistance. Antimicrob Agents Chemother. 2017;61:e02665-16. doi: 10.1128/AAC.02665-16.
101. Venturini C, Hassan KA, Chowdhury PR, Paulsen IT, Walker MJ, et al. Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS26-related molecular signature isolated from different Escherichia coli pathotypes from different hosts. PLoS One. 2013;8:e78862. doi: 10.1371/journal.pone.0078862.
102. McKinnon J, Chowdhury PR, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. Int J Antimicrob Agents. 2018;52:430–435. doi: 10.1016/j.ijantimicag.2018.06.017.
103. McKinnon J, Chowdhury PR, Djordjevic SP. Molecular analysis of an IncF ColV-Like plasmid lineage that carries a complex resistance locus with a trackable genetic signature. Microb Drug Resist. 2020;26:787–793. doi: 10.1089/mdr.2019.0277.
104. Bie L, Fang M, Li Z, Wang M, Xu H. Identification and characterization of new resistance-conferring SGI1s (Salmonella genomic island 1) in Proteus mirabilis . Front Microbiol. 2018;9:03172. doi: 10.3389/fmicb.2018.03172.
105. Alonso CA, González-Barrio D, Tenorio C, Ruiz-Fons F, Torres C. Antimicrobial resistance in faecal Escherichia coli isolates from farmed red deer and wild small mammals. Detection of a multiresistant E. coli producing extended-spectrum beta-lactamase. Comp Immunol Microbiol Infect Dis. 2016;45:34–39. doi: 10.1016/j.cimid.2016.02.003.
106. Mendes Moreira A, Couvé-Deacon E, Bousquet P, Chainier D, Jové T, et al. Proteae: a reservoir of class 2 integrons? J Antimicrob Chemother. 2019;74:1560–1562. doi: 10.1093/jac/dkz079.
107. Mora-Rillo M, Fernández-Romero N, Navarro-San Francisco C, Díez-Sebastián J, Romero-Gómez MP, et al. Impact of virulence genes on sepsis severity and survival in Escherichia coli bacteremia. Virulence. 2015;6:93–100. doi: 10.4161/21505594.2014.991234.
108. Bonacorsi S, Houdouin V, Mariani-Kurkdjian P, Mahjoub-Messai F, Bingen E. Comparative prevalence of virulence factors in Escherichia coli causing urinary tract infection in male infants with and without bacteremia. J Clin Microbiol. 2006;44:1156–1158. doi: 10.1128/JCM.44.3.1156-1158.2006.

109. Lefort A, Panhard X, Clermont O, Woerther P-L, Branger C, et al. Host factors and portal of entry outweigh bacterial determinants to predict the severity of Escherichia coli bacteremia. J Clin Microbiol. 2011;49:777–783. doi: 10.1128/JCM.01902-10.
110. Jauréguy F, Carbonnelle E, Bonacorsi S, Clec'h C, Casassus P, et al. Host and bacterial determinants of initial severity and outcome of Escherichia coli sepsis. Clin Microbiol Infect. 2007;13:854–862. doi: 10.1111/j.1469-0691.2007.01775.x.
111. Brumbaugh AR, Smith SN, Mobley HLT. Immunization with the yersiniabactin receptor, FyuA, protects against pyelonephritis in a murine model of urinary tract infection. Infect Immun. 2013;81:3309–3316. doi: 10.1128/IAI.00470-13.
112. Wright KJ, Hultgren SJ. Sticky fibers and uropathogenesis: bacterial adhesins in the urinary tract. Future Microbiol. 2006;1:75–87. doi: 10.2217/17460913.1.1.75.
113. Miajlovic H, Mac Aogáin M, Collins CJ, Rogers TR, Smith SGJ. Characterization of Escherichia coli bloodstream isolates associated with mortality. J Med Microbiol. 2016;65:71–79. doi: 10.1099/jmm.0.000200.
114. Dale AP, Pandey AK, Hesp RJ, Belogiannis K, Laver JR, et al. Genomes of Escherichia coli bacteraemia isolates originating from urinary tract foci contain more virulence-associated genes than those from non-urinary foci and neutropaenic hosts. J Infect. 2018;77:534–543. doi: 10.1016/j.jinf.2018.10.011.
115. Croxen MA, Finlay BB. Molecular mechanisms of Escherichia coli pathogenicity. Nat Rev Microbiol. 2010;8:26–38. doi: 10.1038/nrmicro2265.
116. Chowdhury PR, McKinnon J, Liu M, Djordjevic SP. Multidrug resistant uropathogenic Escherichia coli ST405 with a novel, composite IS26 transposon in a unique chromosomal location. Front Microbiol. 2018;9:3212. doi: 10.3389/fmicb.2018.03212.
117. Johnson JR, Johnston BD, Porter SB, Clabots C, Bender TL, et al. Rapid emergence, subsidence, and molecular detection of Escherichia coli sequence type 1193-fimH64, a new disseminated multidrug-resistant commensal and extraintestinal pathogen. J Clin Microbiol. 2019;57:e01664-18. doi: 10.1128/JCM.01664-18.
118. Blanco J, Mora A, Mamani R, López C, Blanco M, et al. Four main virotypes among extended-spectrum-β-lactamase-producing isolates of Escherichia coli O25b:H4-B2-ST131: bacterial, epidemiological, and clinical characteristics. J Clin Microbiol. 2013;51:3358–3367. doi: 10.1128/JCM.01555-13.
119. Olesen B, Scheutz F, Andersen RL, Menard M, Boisen N, et al. Enteroaggregative Escherichia coli O78:H10, the cause of an outbreak of urinary tract infection. J Clin Microbiol. 2012;50:3703–3711. doi: 10.1128/JCM.01909-12.
120. Boll EJ, Stegger M, Hasman H, Roer L, Overballe-Petersen S, et al. Emergence of enteroaggregative Escherichia coli within the ST131 lineage as a cause of extraintestinal infections. bioRxiv. 2018;435941
121. Lara FBM, Nery DR, de Oliveira PM, Araujo ML, Carvalho FRQ, et al. Virulence markers and phylogenetic analysis of Escherichia coli strains with hybrid EAEC/UPEC genotypes recovered from sporadic cases of extraintestinal infections. Front Microbiol. 2017;8:146. doi: 10.3389/fmicb.2017.00146.
122. Boll EJ, Struve C, Boisen N, Olesen B, Stahlhut SG, et al. Role of enteroaggregative Escherichia coli virulence factors in uropathogenesis. Infect Immun. 2013;81:1164–1171. doi: 10.1128/IAI.01376-12.

123. Dobrindt U, Chowdary MG, Krumbholz G, Hacker J. Genome dynamics and its impact on evolution of Escherichia coli . Med Microbiol Immunol. 2010;199:145–154. doi: 10.1007/s00430-010-0161-2.
124. Calhau V, Ribeiro G, Mendonça N, Da Silva GJ. Prevalent combination of virulence and plasmidic-encoded resistance in ST 131 Escherichia coli strains. Virulence. 2013;4:726–729. doi: 10.4161/viru.26552.
125. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Nolan LK. Characterizing the APEC pathotype. Vet Res. 2005;36:241–256. doi: 10.1051/vetres:2004057.
126. Johnson TJ, Wannemuehler Y, Johnson SJ, Stell AL, Doetkott C, et al. Comparison of extraintestinal pathogenic Escherichia coli strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. Appl Environ Microbiol. 2008;74:7043–7050. doi: 10.1128/AEM.01395-08.
127. Gérard F, Pradel N, Wu L-F. Bactericidal activity of colicin V is mediated by an inner membrane protein, SdaC, of Escherichia coli . J Bacteriol. 2005;187:1945–1950. doi: 10.1128/JB.187.6.1945-1950.2005.
128. Chattopadhyay S, Weissman SJ, Minin VN, Russo TA, Dykhuizen DE, et al. High frequency of hotspot mutations in core genes of Escherichia coli due to short-term positive selection. Proc Natl Acad Sci USA. 2009;106:12412–12417. doi: 10.1073/pnas.0906217106.
129. Denamur E, Bonacorsi S, Giraud A, Duriez P, Hilali F, et al. High frequency of mutator strains among human uropathogenic Escherichia coli isolates. J Bacteriol. 2002;184:605–609. doi: 10.1128/JB.184.2.605-609.2002.
130. Chen SL, Hung C-S, Xu J, Reigstad CS, Magrini V, et al. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. Proc Natl Acad Sci USA. 2006;103:5977–5982. doi: 10.1073/pnas.0600938103.
131. Nielsen KL, Stegger M, Godfrey PA, Feldgarden M, Andersen PS, et al. Adaptation of Escherichia coli traversing from the faecal environment to the urinary tract. Int J Med Microbiol. 2016;306:595–603. doi: 10.1016/j.ijmm.2016.10.005.
132. Eddy SR. Where did the BLOSUM62 alignment score matrix come from? Nat Biotechnol. 2004;22:1035–1036. doi: 10.1038/nbt0804-1035.
133. Song D, Chen J, Chen G, Li N, Li J, Fan J, Bu D, Li SC. Parameterized BLOSUM Matrices for Protein Alignment. IEEE/ACM Trans Comput Biol Bioinform. 2015 May-Jun;12(3):686-94. doi: 10.1109/TCBB.2014.2366126. PMID: 26357279.
134. Ragheb MN, Thomason MK, Hsu C, Nugent P, Gage J, et al. Inhibiting the evolution of antibiotic resistance. Mol Cell. 2019;73:157–165. doi: 10.1016/j.molcel.2018.10.015.
135. Selby CP. Mfd protein and transcription-repair coupling in Escherichia coli . Photochem Photobiol. 2017;93:280–295. doi: 10.1111/php.12675.
136. Thorpe HA, Bayliss SC, Hurst LD, Feil EJ. Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. Genetics. 2017;206:363–376. doi: 10.1534/genetics.116.195784.
137. Guttman DS, Dykhuizen DE. Clonal divergence in Escherichia coli as a result of recombination, not mutation. Science. 1994;266:1380–1383. doi: 10.1126/science.7973728.

Chapter 6: Genomic comparisons of Escherichia coli ST131 from Australia

6.1. Declaration

Authors:
Dmitriy Li[1], Ethan R. Wyrsch[1], Paarthiphan Elankumaran[1], Monika Dolejska[2,3,4], Marc S. Marenda[5], Glenn F. Browning[5], Rhys N. Bushell[5], Jessica McKinnon[1], Piklu Roy Chowdhury[1], Nola Hitchick[6], Natalie Miller[6], Erica Donner[7], Barbara Drigo[7], Dave Baker[8], Ian G. Charles[8], Timothy Kudinha[9], Veronica M. Jarocki[1] and Steven P. Djordjevic[1]

Author affiliations:
[1] iThree Institute, University of Technology Sydney, Ultimo, NSW, Australia
[2] CEITEC VETUNI, University of Veterinary Sciences Brno, Brno, Czech Republic
[3] Department of Biology and Wildlife Disease, Faculty of Veterinary Hygiene and Ecology, University of Veterinary Sciences Brno, Czech Republic
[4] Biomedical Center, Charles University, Czech Republic
[5] Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Victoria, Australia.
[6] San Pathology, Sydney Adventist Hospital, Wahroonga, NSW 2076, Australia.
[7] Future Industries Institute, University of South Australia, Adelaide, South Australia, Australia
[8] Quadram Institute, Norwich, United Kingdom
[9] Central West Pathology Laboratory, Charles Sturt University, Orange, NSW, 2800, Australia

Published in:
Microbial Genomics, December 2021, doi: 10.1099/mgen.0.000721

Author contribution statement
D.L.: formal analysis, investigation, visualisation, writing – original draft. E.R.W.: data curation, formal analysis. P.E.: sequencing and analysis of canine ST131. E.D. and B.D.; sampling and isolation of environmental ST131 isolates. D.B. and I.G.C.: sequencing of environmental ST131 genomes. P.R.C. and J.M.: sequencing and analysis of hospital ST131 isolates. T.K.: isolation and investigation of hospital ST131 isolates. G.F.B., M.S.M. and R.N.B.: isolation and provision of canine isolates. N.H. and N.M.: isolation and provision of hospital ST131 isolates. M.D.: isolation and microbiological and molecular biological analysis of ST131 isolates from gulls. V.M.J.: conceptualisation, formal analysis, supervision, visualisation, writing – original draft, review and editing and S.P.D.: conceptualisation, supervision, funding acquisition, project administration, writing – original draft, review, and editing.

6.2. Abstract

Escherichia coli ST131 is a globally dispersed ExPEC lineage contributing significantly to hospital and community acquired urinary tract and BSIs. Here we describe a detailed phylogenetic analysis of the whole genome sequences of 284 Australian ST131 *E. coli* isolates from diverse sources, including clinical, food and companion animals, wildlife, and the environment. Our phylogeny and the results of SNP analysis show the typical ST131 clade distribution with clades A, B and C clearly displayed, but no niche associations were observed. Indeed, interspecies relatedness was a feature of this study. Thirty-five isolates (29 of human and six of wild bird

origin) from clade A (32 *fimH*41, 2 *fimH*89, 1 *fimH*141) were observed to differ by an average of 76 SNPs. Forty-five isolates from clade C1 from four sources formed a cluster with an average of 46 SNPs. Within this cluster, human sourced isolates differed by approximately 37SNPs from isolates sourced from canines, approximately 50 SNPs from isolates from wild birds, and approximately 52 SNPs from isolates from wastewater. Many ST131 carried resistance genes to multiple antibiotic classes and while 41 (14%) contained the complete class one integron–integrase *intI1*, 128 (45%) isolates harboured a truncated *intI1* (462–1014bp), highlighting the ongoing evolution of this element. The module *intI1–dfrA17–aadA5–qacEΔ1–sul1*–ORF–*chrA–padR–IS6100–mphR–mrx–mphA*, conferring resistance to trimethoprim, aminoglycosides, quaternary ammonium compounds, sulphonamides, chromate, and macrolides, was the most common structure. Most (73%) Australian ST131 isolates carry at least one ESBL gene, typically $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$. Notably, dual *parC*-1aAB and *gyrA*-1AB fluoroquinolone resistant mutations, a unique feature of clade C ST131 isolates, were identified in some clade A isolates. The results of this study indicate that the ST131 population in Australia carries diverse ARG and plasmid replicons and indicate cross-species movement of ST131 strains across diverse reservoirs.

6.3. Impact Statement

ST131 is a leading ExPEC ST but the major clades A, B and C display different pathobiology. Clade C, in particular C2, is considered the most clinically relevant as it is a frequent cause of urinary tract and BSIs and is resistant to multiple antibiotics, including extended spectrum β-lactams. However, its importance as a global ExPEC lineage is distorted because many studies focus on clinical ESBL-producing ExPEC. Clade A ST131 are poorly studied but in terms of lethality are considered equivalent to clade C in a mouse sepsis model of infection, while clade B ST131 are often zoonotic pathogens with an established poultry link. Here we performed a phylogenetic analysis of 284 Australian isolates of ST131 irrespective of their drug resistance phenotype or source. We identified multiple clusters of highly related ST131 isolates representative of clades A, B and C especially among those from humans, companion animals and synanthropic birds. Evidence of transmission of closely related clade C2 isolates within and between different Australian hospitals was also observed. Notably, clade A isolates carrying $bla_{\text{CTX-M-27}}$ and dual *parC*-1aAB and *gyrA*-1AB fluoroquinolone resistant mutations, a feature previously considered unique to clade C isolates.

6.4. Data summary

The genome sequences of OBH isolates (HOS77–HOS99) have been deposited in GenBank under BioProject PRJNA705804 with SRA accession numbers SRR13843903 to SRR13843925. The

genome sequences of Melbourne Veterinary Collection (MVC) isolates have been deposited in GenBank under the following SRA accession numbers: SRR14629758, SRR14629798, SRR14629727, SRR14629683, SRR14629824, SRR14629814, SRR14629805, SRR14629641, SRR14629609 and SRR14629794. The genome sequences of Australian silver gull isolates (with 'CE' prefix and a single 1716h isolate) have been deposited in GenBank under BioProject PRJNA630096 with SRA accession numbers SRR13834344 to SRR13834353. The genome sequences of South Australia Environment (SAE) isolates are available at GenBank under BioProject PRJNA706036, SRA accession numbers SRR14763817 to SRR14763821. For Sydney Adventist Hospital (SAH) isolates the genome sequence of isolate SAH2009_36 has been published previously under SRA accession number SRX5100115. The genome sequences of all other isolates have been deposited in Genbank under BioProject PRJNA721525 with SRA accession numbers from SRR14229529 to SRR14229538. Assemblies and annotated genomes for all Australian ST131 genomes used in this study are available with the online version of this article.

6.5. Introduction

Escherichia coli ST131, and emerging variants such as ST8196 (1), are significant contributors to hospital- and community-acquired UTIs and BSIs (1–4). *E. coli* ST131 rose to prominence in 2008 and appeared simultaneously on three continents (5). The ST131 lineage belongs to phylogroup B2 and most clinical isolates have an O25b:H4 serotype. *E. coli* ST131 with serotype O16:H5 is also well described, particularly in *fimH*41 isolates that carry the ESBL gene $bla_{\text{CTX-M-14}}$ (6).

The molecular phylogeny of *E. coli* ST131 describes three major sublineages, clades A, B and C, that are primarily delineated based on fimbrial adhesion type 1 (*fimH*) alleles and the carriage of ARGs (7, 8). Clade A (*fimH*41), the most divergent of the three clades, probably emerged in Southeast Asia in the 1880s (9). Recently, clade A isolates have been recovered from wastewater (10) that displayed resistance to drugs used to treat UTIs (fluoroquinolones, gentamicin, sulfamethoxazole-trimethoprim) (11–13). In terms of lethality, clade A isolates with serotype O16:H5 are on a par with O25b:H4 in a mouse sepsis model of infection (6). Australian isolates of ST131 with serotype O16:H5 are linked with men and reproductive-age women suffering with pyelonephritis (14).

Clade B probably emerged in the mid-1900s (9, 15) and carries the most diverse collection of fimbrial adhesion alleles (*fimH*22, H35, H27, H31 and H94) with the *fimH*22 allele predominating in about 65 % of sequenced genomes (9, 16, 17). ST131 clade B isolates are potentially zoonotic as isolates causing human disease have been linked with production of food animals (18),

particularly poultry (19, 20). Clade B isolates often carry ColV plasmids (18, 19), which are linked with the ability of APEC to cause colibacillosis in poultry (21, 22), and contribute to pathogenesis in different animal models (23–25). ColV plasmids carry an important arsenal of VAGs (22) and increasingly carry ARGs (26, 27). ColV plasmid carriage has been linked to the ability of commensal *E. coli* to cause urosepsis (26), and novel *E. coli* STs such as ST301 (28) that have caused recent and severe outbreaks of haemolytic uremic syndrome in Europe and neonatal meningitis (23). Of particular concern is a recent report of MDR *mcr*-carrying *E. coli* ST131 *fimH*22 (clade B) from poultry with colibacillosis in Brazil, the world's largest exporter of poultry products (29).

The results of several studies indicate that ST131 clade C split from clade B in North America between 1985 and 1994 (9, 15, 16, 30) and comprises three subclades: C0, C1 and C2 (31). Clade C0 was established before clinical use of fluoroquinolones became widespread in human and veterinary medicine in the 1980s and is considered ancestral (16). Clades C1 and C2 (almost exclusively *fimH*30) are globally dominant fluoroquinolone-resistant sublineages that carry dual mutations in chromosomal *gyrA* and *parC* genes. Clade C isolates include sublineage C1, defined as H30R and sublineage C2, referred to as H30Rx, which frequently carries the ESBL gene $bla_{\text{CTX-M-15}}$ that encodes resistance to the extended-spectrum cephalosporins used frequently in clinical medicine. Subpopulations of clade C isolates carry $bla_{\text{CTX-M-27}}$ or $bla_{\text{CTX-M-14}}$, but their frequency is considerably less than those that carry $bla_{\text{CTX-M-15}}$ (17). It is posited that the widespread use of fluoroquinolones and extended-spectrum cephalosporins for the past 30 years provided a selection pressure that drove the global expansion of clades C1/H30R and C2/H30Rx (9). ST131 clade C variants carrying genes encoding resistance to carbapenem antibiotics used to treat infections resistant to cephalosporins have also been described (32, 33). Specific IncF-plasmid lineages have been linked to the different clade C lineages with F1:A2:B20 plasmids predominating in clade C1 and F2:A1:B− in clade C2 (31). While F1:A2:B20 and F2:A1:B− plasmids are associated with CTX-M carriage and dissemination (34, 35), F29:A−:B10 plasmids, such as pUTI89, are linked to ExPEC virulence (36).

E. coli ST131 is a significant contributor to UTIs and sepsis in Australia and elsewhere. The COmmunity Onset ESBL and AmpC *E. coli* Study (COOEE Study) identified ST131 to comprise about 8% of community-onset *E. coli* infections in Australia and New Zealand (37, 38). In a large study of urinary tract isolates from reproductive age women (15–45 years of age) from the central west region of NSW, ST131 dominated isolates from patients with pyelonephritis and most also displayed resistance to fluoroquinolones. Resistance to fluoroquinolones was also a feature of ST131 isolates from patients with cystitis (14). The results of a recent WGS study of

67 trimethoprim-resistant ExPEC from patients with UTIs from a major regional hospital in NSW, showed that ST131 represented 28 % of isolates (39) (Chapter 5). Only one ST131 isolate carried a H27 *fimH* allele (clade B) with the remaining isolates carrying a H30 allele (clade C), the majority of which were ESBL $bla_{\text{CTX-M-15}}$-associated H30Rx (clade C2). Another Australian study that assessed 81 *E. coli* isolated from patients with BSIs at Concord Hospital Sydney, identified 15 ST131, most of which were community acquired and MDR (36, 40).

E. coli ST131 has been isolated from pigs (18), poultry (19) companion animals (41) and wild birds (42) in Australia, but WGS has not been performed on a diverse Australian collection – from human, animal, and environmental sources – to examine diversity or interspecies transmission. Here we characterised the genomes of 284 ST131 isolates of Australian origin sourced from humans, companion animals, wastewater, and urban wildlife to garner insights into the ARGs and VAGs they carry. A detailed phylogenetic and pangenome analysis was also undertaken.

6.6. Methods

6.6.1. Genome sequences

This study used both published (*n* = 226) and unpublished (*n* = 58) *E. coli* ST131 genomes. For previously published sequences, EnteroBase (accessed 08.04.2020) was used to screen for all ST131 genomes isolated in Australia with year of isolation available in the associated metadata. All Enterobase sourced isolates have their SRA accession numbers as isolate names. Short-read sequences of these genomes were downloaded using parallel-fastq-dump (github.com/rvalieris/parallel-fastq-dump). All genomes used in this study were assembled using shovill v1.0.4 (github.com/tseemann/shovill) using the SPAdes option and minimal contig length of 200 bp. All genomes underwent and passed quality control using assembly-stats software (github.com/sanger-pathogens/assembly-stats). Assembly statistics for this collection are available in Appendix 17. For unpublished sequences, the isolate sampling and sequencing methods used are described below.

6.6.2. Sampling and WGS of Australian ST131 isolates

Sampling and WGS methods for the 58 ST131 isolates sequenced for this study can be viewed in Appendix 18, which details sampling and WGS methods for MVC isolates, OBH isolates, Australian silver gull isolates, SAE isolates, and SAH isolates.

6.6.3. Phylogeny and SNP analyses

Construction of a maximum-likelihood SNP-based phylogenetic tree and calculation of pairwise SNP distances were executed using the snplord pipeline

(github.com/maxlcummins/pipelord_old/tree/master/snplord), which combines snippy v4.3.6 (github.com/tseemann/snippy) with default settings, recombination filtering software Gubbins v2.3.4 (github.com/sanger-pathogens/gubbins) (43) with default settings, SNP-sites v2.4.1 (github.com/sanger-pathogens/snp-sites) with default settings, and FastTree v2.1.8 (44) using -gtr (generalised time-reversible model) and -nt options. For tree visualisation iToL v4 online based software (45) (itol.embl.de/) was used. Pairwise SNP distance heatmaps were visualised with R v4.0.2 using the packages ggplot2 (github.com/tidyverse/ggplot2) and ggtree (46) (github.com/YuLab-SMU/ggtree). Core genome phylogeny was constructed using IQ-TREE 2 (47) (iqtree.org/) with extended model selection and tree construction by best-fit model (-m MFP) and 1000 bootstrap replicates (-bb 1000) options and using a core genome alignment generated by Roary v3.13.0 (48) (sanger-pathogens.github.io/Roary/) using the fast core gene alignment with MAFFT (-e --mafft) option. FastBAPS grouping (49) was done using R package fastbaps V1.0.4, with default parameters. Third level FastBAPS was used, as it provided separation between ST131 C1 and C2 clades. FastBAPS is a tool for clustering multilocus genotype data as it identifies an approximate fit to a Dirichlet process mixture model (DPM). This is achieved through the Bayesian hierarchical clustering (BHC) algorithm, which efficiently approximates a DPM while ensuring a polynomial time solution (49).

6.6.4. Gene screening

Gene screening was performed as described previously (39). Briefly, serogroups were determined using SerotypeFinder v2.0 (50). FimH typing was performed using FimTyper (50). The ARIBA read-mapping tool with default settings was used to screen for VAGs, ARGs and plasmid replicons using the following reference databases: VirulenceFinder (51), ResFinder (51), PointFinder (52) and PlasmidFinder (52). An additional custom database with AMR-associated class 1 and 2 integrases, and additional ExPEC-associated VAGs was also utilised and can be accessed at githubcom/CJREID/custom_DBs. ARIBA data was processed using a bespoke script accessible at github.com/maxlcummins/pipelord_old/tree/master/aribalord and visualised in R v4.0.2 using package ComplexHeatmap (53) ggplot2 (github.com/tidyverse/ggplot2) and ggtree (46). pMLST screening was performed using pMLST v2.0 (52).

Presence of class 1 integrons was determined using BLAST+ v2.8.1). Contigs harbouring class 1 integrons were further analysed in SnapGene v4.1.9 (snapgene.com) and an integron map was drawn using graphics editing software Krita v4.3.0 (krita.org). Similarly, $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$ genomic context maps were also analysed in SnapGene and drawn in Krita.

6.6.5. Plasmid map visualizations

A BLASTn-based map of pAA-ST131 and other ST131 *fimH*27 *E. coli* as well as pAPEC-O2-ColV were reconstructed using CGView Server (cgview.ca) (54). pUTI89 plasmid mapping was done using BLAST based software ABRicate (github.com/tseemann/abricate) to screen for the plasmid sequence and the map was constructed using an R script available at (github.com/maxlcummins/plasmid_mapR) (21). Isolates with more than 90% pUTI89 plasmid sequence coverage and 90% nucleotide identity were flagged as harbouring a 'pUTI89 like' plasmid. Plasmids were visualized using CGView Server.

6.6.6. Pangenome analysis

For assembly annotation, gene clustering and GWA study, Prokka v1.14.6 (github.com/tseemann/prokka) (55) using --compliant --mincontiglen 200 --genus Escherichia --species coli --gcode 11 options, Roary (sanger-pathogens.github.io/Roary/) (48) and Scoary v1.6.16 (github.com/AdmiralenOla/Scoary) (56) was used with –no_pairwise flag, with 'phenotype' set as subclades were used, respectively. Phandango (jameshadfield.github.io/phandango/) (57) was used for pangenome visualisation.

6.7. Results

6.7.1. Demography

This study added 58 ST131 isolates to a pool of 226 ST131 isolates originating from Australia, and all 284 isolates were analysed here. Isolates from this study were sourced from humans (*n* = 33; 23 OBH isolates, 10 SAH isolates), synanthropic birds (gulls) (*n* = 10, CE isolates), canines (dogs) (*n* = 10, MVC isolates), and wastewater (*n* = 5, SAE isolates). The remaining 226 genomes were acquired from EnteroBase and were derived from human (*n*=183), urban and wild birds (gulls *n* = 36, doves *n* = 3, penguin *n* = 1; total *n* = 40), and poultry, porcine, canine (*n* = 1 each). All isolates, bar the reference genome (EC958), originated from Australia with isolation dates ranging from 2005 to 2019, though the majority were acquired in 2017 (43%, *n* = 122). All associated metadata is available in Appendix 19.

6.7.2. Phylogeny

To determine the evolutionary relationships within this ST131 collection, a maximum-likelihood phylogenetic tree was constructed using the complete genome of *E. coli* strain EC958 (GenBank accession number NZ_HG941718) as a reference. The maximum-likelihood tree was reconstructed using 8200 SNPs from an alignment of 91.2% to the reference genome (4.79 Mbp). The phylogenetic tree clearly shows the typical ST131 clade distribution with clades A, B and C prominently displayed. Subclades C1 and C2 are also distinguishable (Figure 6.1). A C0 subclade was assigned based on the presence of the *fimH*30 allele (15) and the absence of *gyrA*-

1AB and *parC*-1aAB SNPs (2). The predominant serotype across clades B and C was O25b:H4 (*n* = 218, 97 %), while in clade A isolates O16:H5 was dominant (*n* = 51, 86 %).

Recombination filtering reduced the number of SNPs from 8200 to 2377, indicating that 71 % of SNPs were introduced by recombination. Phylogenetic analysis utilizing SNPs without recombination filtering provided a similar topology and clade structure (Appendix 20). In both instances, there was no obvious host-based clustering, with human sourced isolates making up more than 70 % (70 % clade B – 81 % clade C2) of all clades except C0 (100 %, *n* = 4). No gull sourced isolates were present in clades B and C0. Virotyping is another method used to characterise *E. coli* ST131 (58). The method is based on the carriage of a particular set of virulence genes (*afa* FM955459, *iroN*, *ibeA*, *sat*). On the basis of this classification, virotype C (*afa* FM955459−, *iroN*−, *ibeA*−, *sat*+) (71 %, *n* = 201) was the most prevalent followed by virotype A (*afa* FM955459+, *iroN*−, *ibeA*−, *sat*+/−) (15 %, *n* = 42), virotype D (*afa* FM955459−, *iroN*+/−, *ibeA*+, *sat*+/−) (6 %, *n* = 16) and virotype B (*afa* FM955459−, *iroN*+, *ibeA*−, sat+/−) (1 %, *n* = 4). A total of 21 isolates (7 %) could not be classified using this classification scheme. Virotypes demonstrated some cohesion with phylogeny. For example, all clade B isolates were virotype D and this virotype was restricted to these isolates, virotype B is only observed in clade C, and virotype C was most associated with clade C1 isolates.

The total pangenome of 284 Australian ST131 isolates identified 15,050 genes (3453 core genes (>=99 %) 400 soft core genes (>=95 %), 1564 shell genes (>=15 %) and 9,633 cloud genes (<15 %), with 3,558 singletons). A gene presence/absence matrix and a GWA analysis can be viewed in Appendix 21.

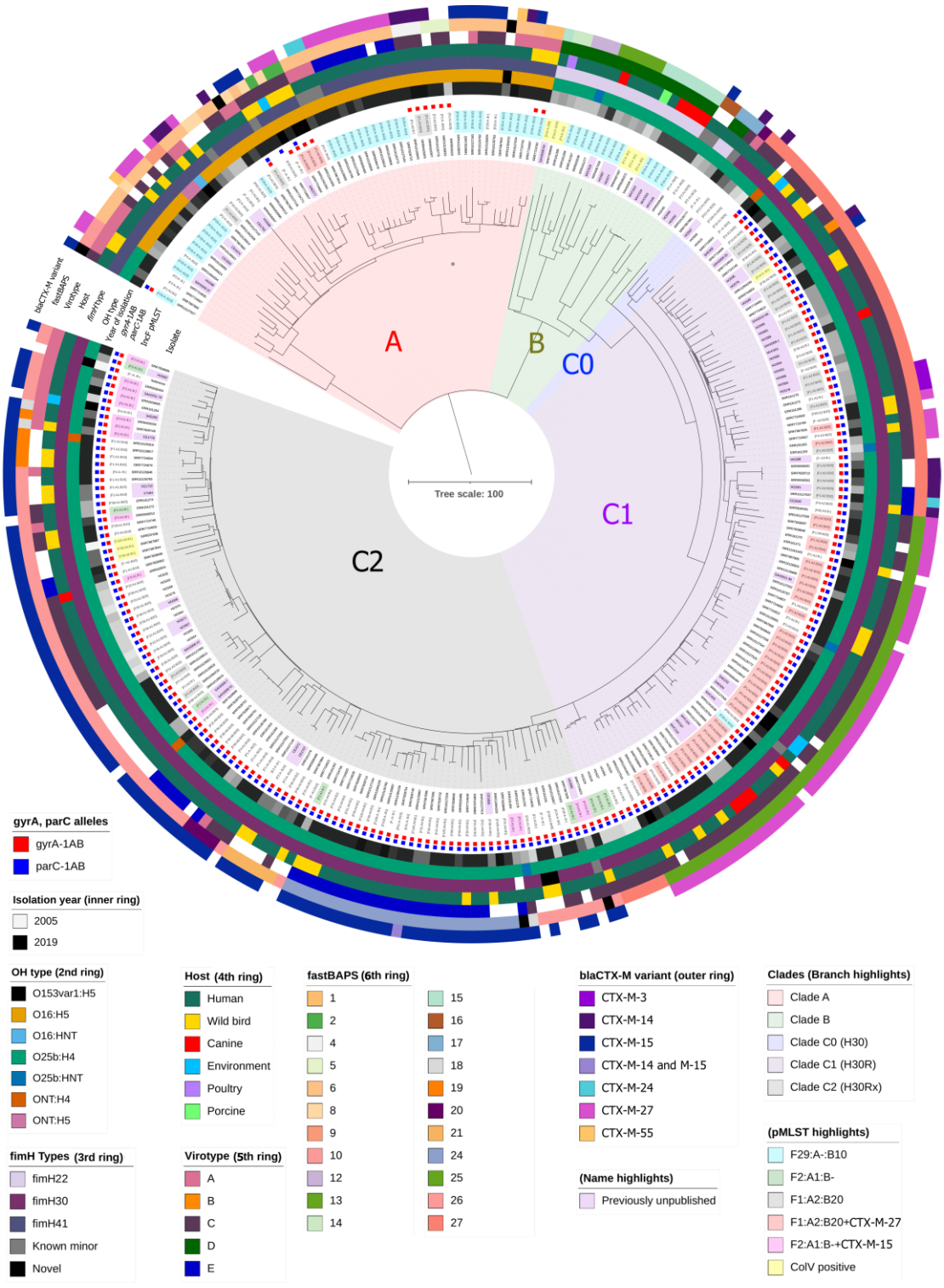

blaCTX-M variant
fastBAPS
Virotype
Host
fimHtype
OH type
Year of isolation
gyrA-1AB
parC1AB
IncF pMLST
Isolate
A
B
C0
C1
C2
Tree scale: 100
gyrA, parC alleles
gyrA-1AB
parC-1AB
Isolation year (inner ring)
2005
2019
OH type (2nd ring)
O153var1:H5
O16:H5
O16:HNT
O25b:H4
O25b:HNT
ONT:H4
ONT:H5
Host (4th ring)
Human
Wild bird
Canine
Environment
Poultry
Porcine
fastBAPS (6th ring)
1
2
4
5
6
8
9
10
12
13
14
15
16
17
18
19
20
21
24
25
26
27
blaCTX-M variant (outer ring)
CTX-M-3
CTX-M-14
CTX-M-15
CTX-M-14 and M-15
CTX-M-24
CTX-M-27
CTX-M-55
Clades (Branch highlights)
Clade A
Clade B
Clade C0 (H30)
Clade C1 (H30R)
Clade C2 (H30Rx)
fimH Types (3rd ring)
fimH22
fimH30
fimH41
Known minor
Novel
Virotype (5th ring)
A
B
C
D
E
(Name highlights)
Previously unpublished
(pMLST highlights)
F29:A-:B10
F2:A1:B-
F1:A2:B20
F1:A2:B20+CTX-M-27
F2:A1:B-+CTX-M-15
ColV positive

Figure 6.1: Maximum-likelihood SNP-based phylogenetic tree of Australian ST131. Inferred using FastTree2, EC958 as a reference, and recombination filtering. Red and blue squares next to labels, denote presence of the characteristic fluoroquinolone resistance SNPs: *gyrA*-1AB (red) and *parC*-1aAB (blue). Coloured rings represent isolation dates (inner ring; colour gradient from light (2005) to dark (2019), serotype (second ring from centre), *fimH* type of isolate (third ring), host (fourth ring), virotype (fifth ring), fastBAPS groups (sixth ring) and $bla_{CTX\text{-}M}$ variant (outer ring). Sections of the tree are coloured according to subclade (red = A, green = B, blue = C0, violet = C1 and black = C2). Inner tip label denotes isolate name, with names highlighted in purple representing previously unpublished genomes. Outer tip labels show which IncF plasmids are present in each isolate, with pMLSTs important for *E. coli* ST131 evolution highlighted. Tree scale bars represent the number of substitutions per site of alignment.

6.7.3. Pairwise SNP distance comparisons

Pairwise SNP distance heatmaps were produced to better illustrate the unique differences between A, B, and C clades. Three types of heatmaps were constructed, one without recombination filtering, one with recombination filtering, and one covering SNPs due to recombination only (Appendix 22). The comparisons of pairwise SNP distances between recombination filtered and unfiltered heatmaps clearly show that differences between and within clades is largely due to recombination, with the number of SNPs reduced approximately 25 times post-filtering (Table 6.1). On the basis of this observation the heatmap of recombination only SNPs was constructed, and it was observed that within clades, clade C isolates had a low number of recombination only SNPs with an average of 58 (range: 0–1487), while clade B isolates exhibited higher recombination SNP variation with an average of 485 (0–1324) SNPs. Clade A had an average recombination SNP count of 383 (0–1505), however 35 out of the 59 isolates displayed a low recombination SNP count 23 (0–205) on average.

Table 6.1: Average SNP count difference between five clades. Numbers are average of recombination filtered/unfiltered

	Clade A				
Clade A	32/415	Clade B			
Clade B	136/3531	37/558	Clade C0		
Clade C0	125/3639	65/1670	24/420	Clade C1	
Clade C1	128/3548	70/1534	35/249	22/33	Clade C2
Clade C2	131/3560	73/1545	37/283	35/87	23/104

6.7.4. Closely related clade clusters

Our SNP analyses identified three clusters of closely related isolates from three clades - clade A, C1 and C2. It is important to identify potential for interspecies transmission and transfer from a public health perspective (59). Thus, each closely related cluster was investigated further with

additional SNP analyses performed using clade specific references and recombination filtering to provide greater SNP resolution.

Thirty-five isolates from clade A (32 *fimH*41, 2 *fimH*89, 1 *fimH*141) were observed to differ by an average of 76 SNPs (Figure 6.2a). These isolates originated from 29 human and six synanthropic bird samples. Isolates from synanthropic bird samples averaged 77 SNPs from human sourced isolates, with the smallest SNP count occurring between CE1674 (silver gull) with human sourced SRR10126843 at 45 SNPs. The average SNP distance between human sourced isolates was 76, however the cluster also contained nine human sourced isolates (two from 2015, seven from 2017) that differed by an average of 22 SNPs (0–31). The average SNP distance between synanthropic bird sourced isolates was 69 (23-88).

A set of 45 isolates from clade C1 from four sources (canine, wastewater, human, synanthropic bird) formed a cluster with an average of 46 (0–70) SNPs (Figure 6.2b). The most closely related non-human sourced isolates to human-sourced isolates originated from canines (average 37 SNPs), followed by synanthropic birds (average 50 SNPs), and then wastewater (average 52 SNPs). The cross-host isolates with the lowest SNP counts were SRR7724766 (bird) and SRR7828694 (human) at 24 SNPs, SRR11341638 (bird) and MVC405 (canine) at 27 SNPs, and MVC405 (canine) and two human sourced isolates SRR5936484 and ERR161271 at 23 and 24 SNPs, respectively. Across human sourced isolates the average SNP count was 47 (3–68).

The third closely related cluster consisted of 13 isolates originating from three geographically distant hospitals – OB (rural NSW), SAH (urban NSW), and Princess Alexandra Hospital Brisbane (urban Queensland) – with an average SNP count of 19 (0–69) (Figure 6.2c). All OBH isolates ('HOS' isolates) were exceptionally close at an average of 6 (0–19) SNPs despite being isolated from different patients and pathologies (seven patients with cystitis and four patients with pyelonephritis). Additionally, there were average SNP counts of 44 (41–54) SNPs from SAH isolate (SAH2009_47) and 59 (55–69) SNPs from Princess Alexandra Hospital isolate (ERR161252).

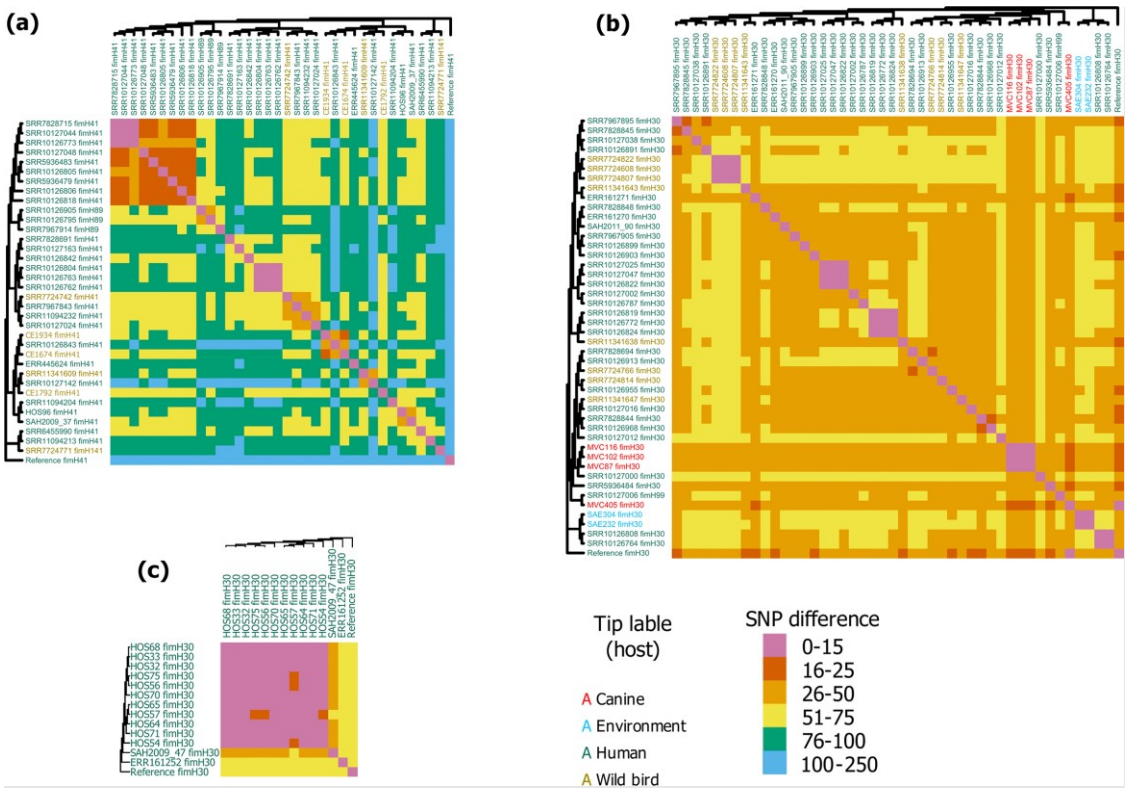


Figure 6.2: Closely related clade clusters. a) Heatmap illustrating pairwise SNP distances for 35 isolates in clade A. Both axes have a maximum-likelihood phylogenetic tree presented as a cladogram, constructed using these 35 isolates and clade A reference strain SE15 (ASM1048v1) (98% genome coverage; 4.73 Mbp; 970 SNPs total). (b) Heatmap illustrating pairwise SNP distances for 45 isolates from clade C1. Both axes have a maximum-likelihood phylogenetic tree presented as a cladogram, constructed using these 45 isolates and clade C1 reference strain 81 009 (ASM215724v1) (97% genome coverage; 4.98 Mbp; 711 SNPs total). (c) Heatmap showing SNP distances of 13 closely related hospital isolates from subclade C2. Both axes have a maximum-likelihood phylogenetic tree presented as a cladogram, constructed using only these 13 isolates, with EC958 as the reference strain (96% genome coverage; 5.02 Mbp; 133 SNPs total). All trees were constructed using recombination-filtered SNP alignments.

6.8. Gene screening

This collection of 284 isolates was screened for the presence of VAGs, ARGs, plasmid replicons and IncF pMLSTs (Figure 6.3, Appendix 19). Details are provided in the subsequent subsections.

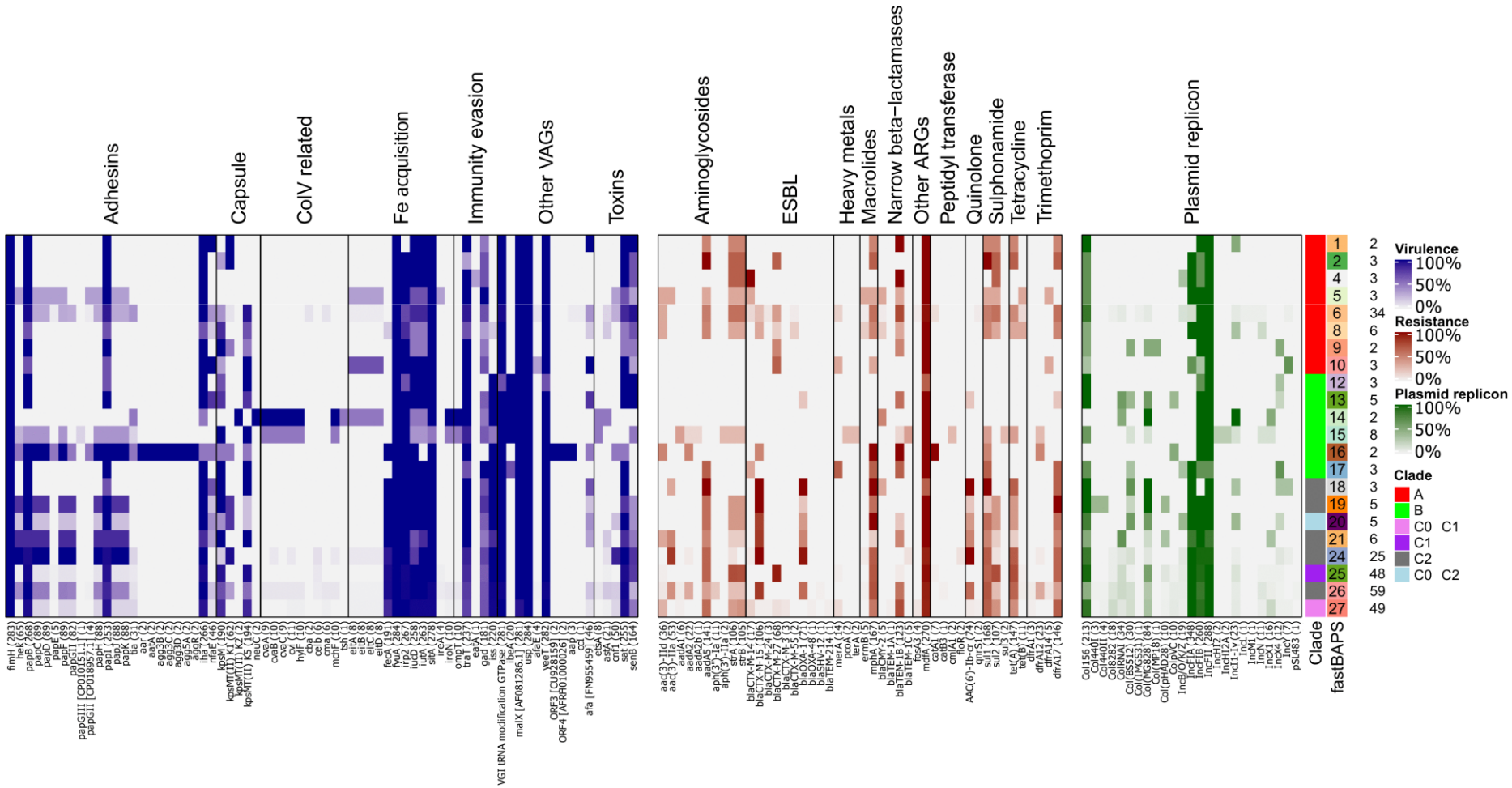


Figure 6.3: ARG, VAG and IncF pMLST abundances in Australian ST131 population. Heatmap based on clustering by fastBAPS (fastBAPS groups consisting of singular isolates omitted). Relative gene abundance presented as blue (VAGs), red (ARGs) and green (plasmid replicons) colour gradients ranging from 0% (no colour) to 100% (most opaque). Total numbers of isolates carrying a gene are shown in parentheses after the gene name.

6.8.1. AMR

A total of 44 ARGs were found in the Australian ST131 isolates (Figure 6.3; Appendix 19). Clade B isolates carried the fewest ARGs (Figure 6.4a), with the average isolate possessing genes conferring resistance to only one antibiotic class (Figure 6.4b) while the members of clade C2 were the most resistant, carrying on average ARGs conferring resistance to seven antibiotic classes. Human and gull sourced isolates presented similar numbers of ARGs carried, while canine sourced isolates had comparably lower carriage (Figure 6.4c, d). At least one ESBL gene was present in 73 % (*n*=206) of all isolates. Of the ten ESBL genes identified, *bla*$_{CTX\text{-}M\text{-}15}$ was most prevalent overall (*n*=106) followed by *bla*$_{CTX\text{-}M\text{-}27}$ (*n*=68). *bla*$_{CTX\text{-}M\text{-}27}$ was only present in clades A and C1, while *bla*$_{CTX\text{-}M\text{-}15}$ was most prominent in C2. Notably, a slightly higher percentage of clade A isolates (71%) carried ESBLs than did clade C1 isolates (65 %).

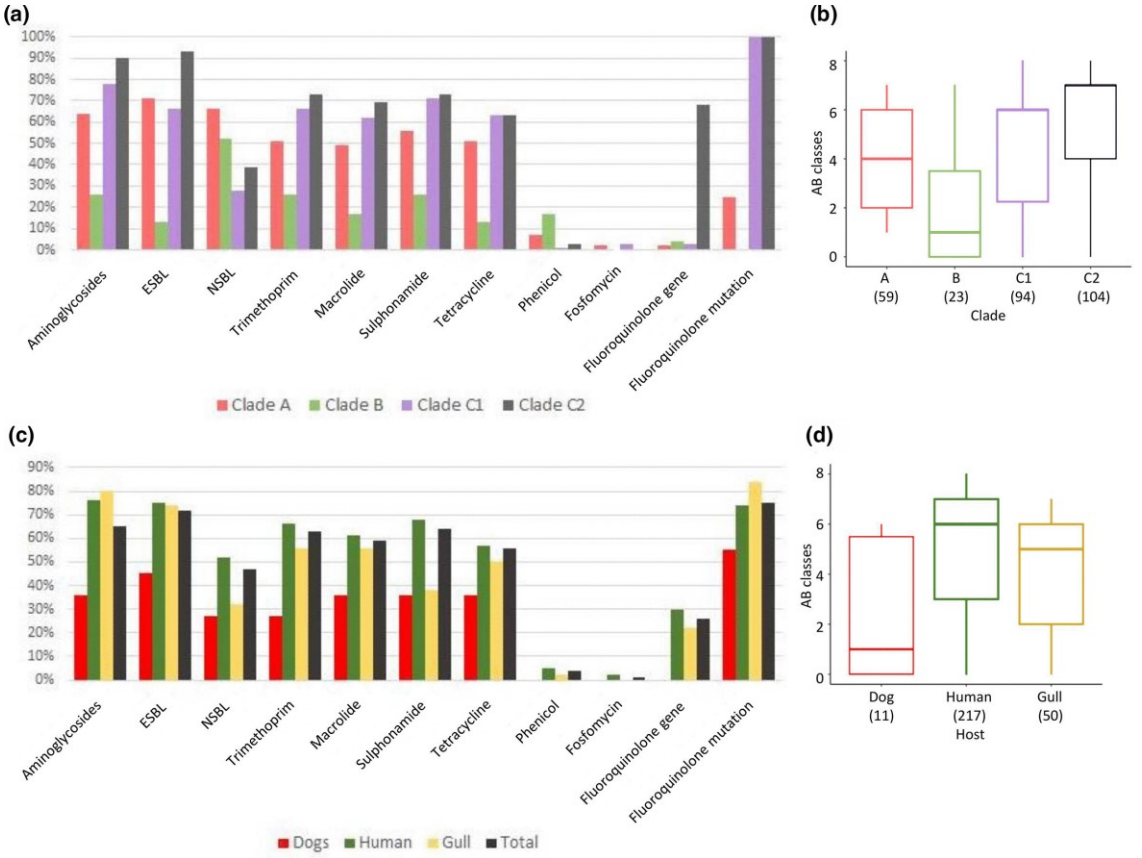


Figure 6.4: Percentage of Australian ST131 isolates carrying ARGs conferring resistance to multiple antibiotics. a) Percentage of isolates carrying one or more ARGs conferring resistance to various antibiotic classes across clades. Clade C0 isolates omitted due to the limited number (n = 4). With Y axis represent percentage of isolates and X axis represent classes of antibiotics. **b)** Box and whisker plot illustrating distribution of clade isolates carrying ARGs conferring resistance to antibiotic (AB) classes. Y axis represent number of antibiotic classes in isolates and X axis ST131 clades. **c)** Percentage of isolates carrying more than one ARG conferring resistance to various antibiotic class across hosts and in total. Environmental (n = 5), porcine (n = 1) and poultry (n = 1) isolates omitted due to the limited numbers. Fluroquinolone mutation refers to the presence of *parC* and *gyrA* mutations. **d)** Box and whisker plot illustrating distribution of host

isolates carrying ARGs conferring resistance to AB classes. The presence of one or more genes counted as conferring resistance. Numbers in brackets are numbers of isolates.

Clade C1 isolates carrying *bla*$_{CTX\text{-}M\text{-}27}$ (C1-M27) reportedly carry a specific phage-like region-M27PP1 (60). We screened for the presence of M27PP1 (>98 % identity and >98 % coverage) and found it in 49 isolates (17 %) (Appendix 23), including three C1 isolates and one C2 isolate that do not possess *bla*$_{CTX\text{-}M\text{-}27}$ and 45 (92 %) C1 isolates carrying *bla*$_{CTX\text{-}M\text{-}27}$. No clade A isolate, *bla*$_{CTX\text{-}M\text{-}27}$ positive or negative, harboured M27PP1. The most common NSBL gene was *bla*$_{TEM\text{-}1B}$ across all hosts (*n* = 123). *bla*$_{OXA\text{-}1}$ was heavily associated with clade C2 isolates (99 %, *n*=70 out of 71).

Genes conferring resistance to aminoglycosides, macrolides, sulphonamide, trimethoprim and tetracyclines were also prevalent among Australian ST131 isolates. Of the ten aminoglycoside resistance genes identified, the most common were *aadA5* (50%, *n* = 141) followed by *strA* (37 %, *n* = 106) and *strB* (37 %, *n* = 105). Macrolide resistance gene *mphA* was found in 59 % of isolates (*n* = 167), while *ermB* was only identified in two gull and three human sourced isolates. Sulphonamide resistance gene *sul1* was also identified in 59 % of isolates (*n* = 168) while *sul2* was identified in 38 % (*n* = 107). *sul3* was only seen in the single porcine sourced isolate and one human sourced isolate, and these two isolates differed by 10 SNPs, as previously described (18). Trimethoprim resistance gene *dfrA17* was the most common of this class and was the only trimethoprim resistance gene identified in gull (56 %, *n* = 28), environmental (80 %, *n* = 4), and canine (27 %, *n* = 3) sourced isolates. While *dfrA17* was also the most abundant in human sourced isolates (51 %, *n* = 111), other trimethoprim resistance genes were also identified in this cohort including *dfrA12* (10 %, *n* = 22), *dfrA14* (2 %, *n* = 5), *dfrA1* (1 %, *n* = 3), *dfrA7* (*n* = 1). Tetracycline resistance gene *tetA* was identified in 52 % of all isolates (*n* = 147) while *tetB* was only present in 11 (4 %) isolates, and only those originating from gulls and humans.

Conversely, genes conferring resistance to phenicol, fluroquinolone and fosfomycin were less prevalent in Australian ST131 isolates. These genes included phenicol resistance genes *catA1* (2 %, *n* = 7), *catB3* (*n* = 1), *cmlA1* (1 %, *n* = 2) and *floR* (1 %, *n* = 2); and fosfomycin resistance gene *fosA3* (1 %, *n* = 4) and fluroquinolone resistance gene *qnrS1* (1 %, *n* = 2). The aminoglycoside gene with extended fluoroquinolone resistance *acc(6')-Ib-cr*, was almost exclusively found in clade C2 isolates (71 out of 74; 96 %). Genes conferring resistance to phenicol and fluroquinolones occurred only in human and gull sourced isolates, and to fosfomycin only in human sourced isolates. Only one Australian ST131 isolate carried a carbapenem resistance gene (*bla*$_{OXA\text{-}48}$) and this isolate was of clade C1 and human sourced.

Regarding mutations conferring resistance to fluoroquinolone, 74% (*n* = 212) of ST131 *E. coli* isolates had *gyrA*-1AB mutations and 71% (*n* = 202) had *parC*-1aAB mutations with 71% (*n* = 201) isolates having both mutations. Clade C accounted for 198 isolates with double gene mutations, the remaining three belonged to clade A isolates. Some uncommon mutations were also identified, including *gyrA*-1AD (one clade A seagull sourced isolate), *parC* (E84K) (one clade B human sourced isolate) and *parC* (S57T) (one clade C human sourced isolate). Of these uncommon mutations, only *gyrA*-1AD has previously been shown to confer fluoroquinolone resistance (61).

6.8.2. Class 1 and 2 integrons and $bla_{\text{CTX-M-15/27}}$ genetic contexts

Class 1 integrons play a central role in AMR because they can capture and express diverse resistance genes and their presence is considered a reliable proxy for an MDR genotype (62). In Australian ST131 isolates, full-length class 1 integron-integrase *intI1* was present in 41 isolates (14%), and a class 2 integron-integrase *intI2* was identified in one isolate. Notably, truncated *intI1* genes were also detected in 128 isolates (45%) with *ΔintI1* ranging from 462 to 1014 bp. These truncations serve as potential epidemiological markers to track MDR ST131 and the mobile elements they carry. For example, a 926 bp truncation was only present in 12 related isolates.

Six primary class 1 integron structures were observed (Figure 6.5a), the most abundant being *intI1*–*dfA17* (trimethoprim resistance)–*aadA5* (aminoglycoside resistance)–*qacEΔ1* (quaternary ammonium compounds resistance)–*sul1* (sulphonamide resistance)–ORF–*chrA* (chromate resistance)–*padR*–IS*6100*–*mphR*–*mrx*–*mphA* (macrolide resistance) found in 47% (*n* = 101) of human, 50% (*n* = 25) of seagull, 27% (*n* = 3) of canine and 80% (*n* = 4) of environment isolates (structure on the black line in Figure 6.5a). Six integron structures carried the mercury resistance transposon Tn*As3* (structure on the pink line in Figure 6.5a) found in clade B (*n* = 1) and clade C1 (*n* = 5) human sourced isolates and two structures (orange line) carried Tn*3* transposon with $bla_{\text{TEM-1B}}$, found in one clade A and B human sourced isolates. Only one isolate (SRR10126957) contained a class 2 integron structure (*intI2*–*dfrA1*–*sat2*–*aadA1*) carried by a clade C2 human sourced isolate from 2017.

Regarding the genetic context of $bla_{\text{CTX-M-15/27}}$ genes, the majority of $bla_{\text{CTX-M-15}}$ genes were located adjacent to IS*Ecp1* (92%, *n*=97) and downstream of Tn2 (91%, *n*=96) (Figure 6.5b). Notably, there were 15 isolates in which the $bla_{\text{CTX-M-15}}$ gene was localised to the chromosome (2 in clade A, and 13 in clade C2 isolates). Concerning the genetic context of $bla_{\text{CTX-M-27}}$, in all examples the gene was situated adjacent to IS*Ecp1* and IS*903B* (Figure 6.5b). Nine clade A

isolates had this insertion within the chromosome, one of which was situated 4219 bp upstream from the type I restriction modification system gene and eight isolates had insertions near a general secretion pathway (*gsp*) operon. These insertion sites serve as epidemiological markers in future ST131 studies. The remaining 59 isolates had *bla*$_{CTX-M-27}$ situated on small contigs with IS*26* situated at contig breaks, and too few genes to ascertain their genetic context. Three clade A isolates carried an identical IS*Ecp1*-*bla*$_{CTX-M-27}$–IS*903B* structure seen in the majority of C1 isolates (Figure 6.5c).

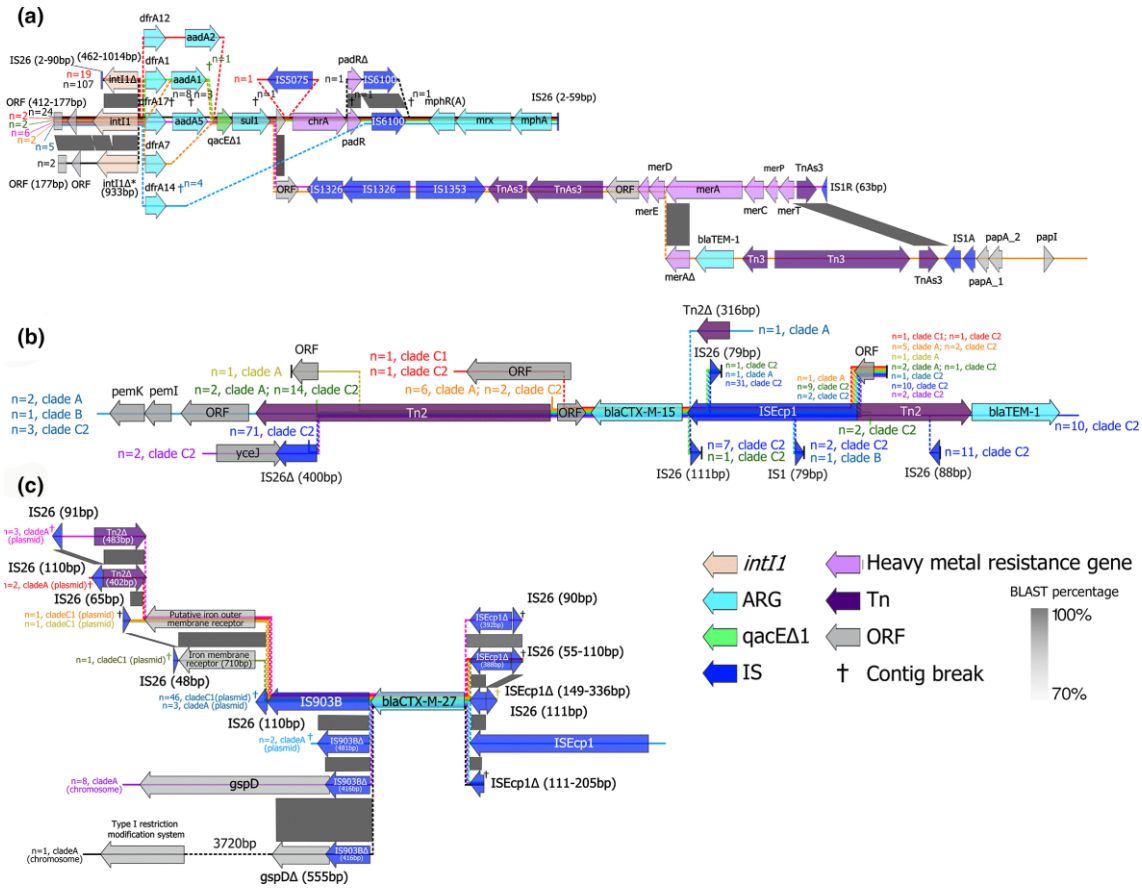


Figure 6.5: Class 1 integron structures and *bla*$_{CTX-M-15}$, *bla*$_{CTX-M-27}$ genomic contexts. a) Schematic map of class one integrons present in Australian ST131 isolates. **b)** Schematic map of genomic context of *bla*$_{CTX-M-27}$. **c)** Schematic map of genomic context of *bla*$_{CTX-M-15}$. In each panel, solid-coloured lines represent individual structures with dashed lines representing variations and connections between structures. Contig breaks denoted by '†'.

6.8.3. VAGs

A total of 72 VAGs were identified in the Australian ST131 isolates (Figure 6.3, Appendix 19). Present in all isolates were ferric yersiniabactin uptake receptor *fyuA* and uropathogenic specific protein *usp*. Other key VAGs featured here were the Sit ferrous iron utilisation system gene *sitA* (98 %, *n* = 258), aerobactin genes *iutA* (93 %, *n* = 263) and *iucD* (91 %, *n* = 258), adhesin *iha* (94 %, *n* = 266), secreted autotransporter toxin *sat* (90 %, *n* = 255) and group two capsule gene *kpsMT*-

II (91%, *n* = 258). The *kpsMT-II* K5 allele was the most common across all Australian ST131 isolates (*n* = 194) and dominated all clades and hosts except for clade C2 isolates, wherein 50% (*n* = 52) carried the K1 allele while 42% (*n* = 44) carried K5. Only two isolates carried *kpsMT-II* K2 (both clade B). The major components of the pyelonephritis-associated pili operon (*papBCDFGHIJK*) were present in 87 isolates (31%), however isolates contained varying degrees of *pap* gene representations, ranging from 94% of isolates containing *papB* (*n* = 268) to 2% carrying *papE* (*n* = 5). In terms of average total number of VAGs, clade C2 isolates had the highest (*n* = 25) followed by clade B isolates (*n* = 23), though clade B isolates contained several unique VAGs including those associated with EAEC such as *agg3BCD*, *agg5A* and *aggR*, as well as those associated with NMEC such as *ibeA* and *neuC*.

6.8.4. Plasmid replicons, IncF pMLSTs and virulence plasmid comparisons

A total of 13 plasmid Inc groups were identified in Australian ST131 isolates (Figure 6.3, Appendix 19), the most abundant being IncFII (96%, *n* = 273) followed by IncFIB (85%, *n* = 242) and IncFIA (71%, *n* = 202). Other Inc groups detected were IncI1 (8%, *n* = 23), IncB/O/K/Z (7%, *n* = 19), IncX4 (4%, *n* = 12), IncX1, IncN (each 4%, *n* = 11), IncY (2%, *n* = 7), IncHI2, IncHI2A (each 1%, *n* = 2), IncL and IncM1 (each *n* = 1).

Among the IncF plasmids a total of 67 pMLSTs were identified, the most common being F1:A2:B20 (25%, *n* = 70), F29:A−:B10 (16%, *n* = 44) and F2:A1:B− (8%, *n* = 23). The dominant pMLST type in clades A and B was F29:A−:B10 (both 53%), F22:A1:B20 in clade C0 (75%), F1:A2:B20 in clade C1 (65%) and F2:A1:B− in clade C2 (19%). F1:A2:B20 and F2:A1:B- are well described and dominant F plasmids in ST131 clade C sublineages (31).

F29:A−:B10 is an important IncF pMLST found in several pandemic, but pan-susceptible, ExPEC lineages (63). F29:A-:B10 is also known to be pMLST of the pUTI89 virulence plasmid (36). F29:A-B10 was the dominant IncF pMLST in Australian clade A and B ST131 isolates. Using pUTI89 (NC_007941.1) as a reference we identified 46 isolates that potentially carry pUTI89-like plasmids (Figure 6.6a). Notably 49% of clade A isolates and 48% of clade B isolates potentially carry a pUTI89-like plasmid, while conversely only 3% of any clade C isolates did. Notably, 13 isolates carrying pUTI89-like plasmids also carried $bla_{\text{CTX-M-27}}$, nine of which were localised to the chromosome (all clade A) and the genetic context for the remainder could not be determined due to small contig size. Similarly, 11 pUTI89-like plasmid carriers also harboured $bla_{\text{CTX-M-15}}$, five of which were chromosomal and the remainder at indeterminable locales. It was also noted that pUTI89-like cargo was associated with fewer plasmid replicons (average 0.7, (0–4) vs average 1.9 (0–6), Wilcoxon Test P=9.6×10−12) in ST131.

ColV plasmids are also associated with ExPEC virulence (24). On the basis of the criteria of Liu *et al.* (19) ten ColV-positive isolates were identified (six in clade B, three in clade C2, one in clade C1), consisting of four types: *iuc* (aerobactin iron acquisition system) and *ets* (*E. coli* transport system) operon-positive, *eit* (ABC iron transport system)-negative (*n*=6); *ets*-positive, *iuc*- and *eit*-negative (*n* = 2); *iuc*-positive, *ets*- and *eit*-negative (*n*=1); *iuc*- and *eit*-positive, *ets*-negative (*n* = 1). Using pAPEC-O2-ColV (NC_007675) as a reference, all ten putative ColV plasmids contained *cvaABC* (colicin V), *iroBCDEN* (salmochelin iron acquisition system) and *sitABCD* (ABC iron transport system) regions (Figure 6.6b).

Two *fimH*27 isolates (clade B) were found harbouring AAF genes (*agg3BCD*, *agg5A*) known to be carried on virulence plasmid pAA, a defining feature of EAEC (64). A hybrid ST131 ExPEC/EAEC lineage carrying pAA caused an outbreak in Denmark (1998 to 2000) (65). While a phylogenetic analysis demonstrated that the two Australian isolates were not closely related to any of the Denmark outbreak strains (Appendix 24), a BLASTn comparison of pAA-ST131 (KY706108; plasmid from a Danish outbreak strain) illustrated that they both potentially carry similar plasmids (Figure 6.6c).

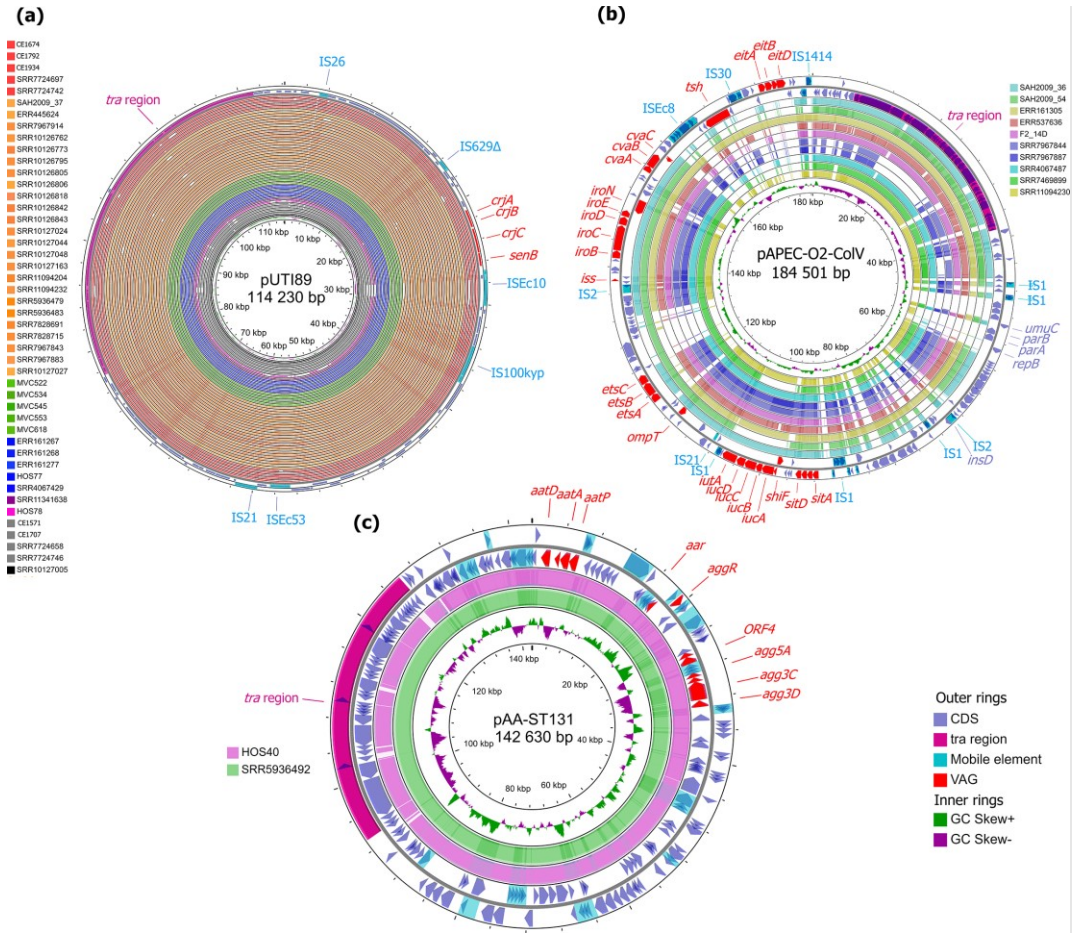


Figure 6.6: BLAST comparisons of three virulence plasmids using CGview. a) Comparisons of pUTI89-like plasmids present in 43 Australian ST131 isolates. Rings coloured by isolate source and clade: red = wild bird, clade A; orange = human, clade A; green = canine, clade B; blue = human, clade B; purple = wild bird, clade C1; pink = human, clade C1; grey = wild bird, clade C2; black = human, clade C2. **b)** Comparisons of pAPEC-O2-ColV-like plasmids present in ten isolates **c)** Comparisons of pAA-like plasmids present in two isolates. VAGs are coloured in red, mobile elements in blue, tra region in scarlet, inner circle show GC skew.

6.9. Discussion

The *E. coli* ST131 lineage was first reported in Australia in 2010 in a study that collected *E. coli* samples from six microbiology laboratories between 2008 and 2009 and from a total of 15 ST131 isolates identified, four produced CTX-M-15 (66). Here, approximately a decade later, we added 58 ST131 genomes to a total pool of 284 ST131 WGS of Australian origin sourced from humans, animals, and the environment. Through genomic analyses several key observations on the Australian ST131 population were made including (1) no clear indication that ST131 clades are niche associated; (2) ESBL genes $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$ dominated; (3) most clade A isolates carried ESBL genes (71%), some of which harboured chromosomal $bla_{\text{CTX-M-27}}$ and others carried dual *parC*-1aAB and *gyrA*-1AB mutations conferring ciprofloxacin resistance, and (4) there was a greater diversity and distribution of F plasmids pMLSTs across clades than previously noted.

Research has emerged indicating that ST131 clades are niche associated, specifically that clade C has adapted to human hosts (3), clade B is zoonotic and dominates meat, pigs (18, 67) and poultry (19), and that clade A is predominantly environmental (10). Others claim genetic commonality and a broad distribution across ecological niches and reservoirs (68, 69). While cautious to extrapolate from a dataset heavily skewed towards human sourced ST131 isolates (76%), our results aligned more with the latter argument of no clear clade orientated niches. Each clade (A, B, C1 and C2) was comprised of approximately ¾ human sourced and approximately ¼ non-human sourced isolates and our phylogenetic analyses did not demonstrate any distinct clustering of these non-human isolates within any clade. Instead, the analysis demonstrated canine, porcine, poultry, synanthropic bird (silver gull) and wastewater sourced isolates dispersed among the human sourced, indicating cross-species transmission. Indeed, our SNP analyses identified clusters of highly related cross-species isolates, with the smallest SNP distance observed between human and silver gull sourced clade C1 isolates (24 SNPs). Clade C1 was found to be particularly clonal with an average of 22 core SNPs. Examples of small cross-species SNP counts in ST131 isolates have been described previously, including an Australian clade B porcine and clinical ST131 isolate separated by 22 SNPs (18) and a clade C1 cat urine (Australia) and human faecal (Spain) ST131 isolate separated by 66 SNPs (41). It is worth mentioning that in the context of hospital outbreaks, a commonly used threshold is often

set at 17 SNPs (70). Experimental studies have estimated the mutation rate to be around 1*10^-3 per genome per generation (71), while household sampling studies suggest a rate of approximately 1.1 mutations per genome per year (72). However, it is important to note that the SNP numbers in the present study could be even lower than these estimates, as pairwise alignments were not performed.

When arguing genetic commonality, recombination should be addressed (73). Recombination facilitates the integration of exogenous DNA, including MGEs, thereby promoting and accelerating adaptive diversification in bacteria (74–76). The recombination rate of commensal *E. coli* is estimated to be relatively low, approximately 0.7 recombination per point mutation (77). Intriguingly, in closely related species such as *Pseudomonas syringae*, this number reach 1.5 (78), while in Salmonella enterica, it can be as remarkable as 30.2 (79). However, unlike commensal *E. coli* recombinations rates in ExPEC are particularly high (75), with as much as 77% of SNPs have been attributed to recombination in ExPEC ST131 (7). The high rates of recombination in ExPEC are thought to be related to a versatility in infection sites (74) This high percentage is congruent with our own 71% of SNPs due to recombination identified. Previous studies have shown that recombination filtering is required to elucidate true ST131 tree topology (16) and indeed in this study only through recombination filtering was the C0 clade made evident. Nevertheless, without recombination filtering, the topology of our phylogenetic tree remained largely unchanged with non-human sourced isolates distributed among human sourced isolates, and closely related cross-species clusters remained intact. This indicates that while recombination contributes considerably to ST131 diversity, it does not alter the apparent cross-species movement within clades.

The *E. coli* ST131 lineage is notorious for having played a central role in the worldwide increase of ESBL-producing *Enterobacteriaceae* which was driven by the capture of $bla_{\text{CTX-M}}$ genes from the environment (80). Copious studies have touted the MDR profiles of ST131 isolates worldwide (2, 3, 81) and the Australian ST131 population appears no different. While only a single Australian ST131 isolate carried a carbapenemase and none carried colistin resistance genes, on average ST131 isolates from Australia carried genes resistant to five antibiotic classes, ranging from clade B isolates averaging resistance to one class, to C2 isolates averaging resistance to seven classes. C2 isolates carried more class one integrons, which are considered reliable indicators for a MDR genotype (62) although notably many of the integrase genes were truncated and therefore are unlikely to be detected with standard PCR primer sets. Overall, the most common structure carrying class 1 integron found in Australian ST131 isolates was *intI1*–*dfrA17*–*aadA5*–*qacEΔ1*–*sul1*–ORF–*chrA*–*padR*–IS*6100*–*mphR*–*mrx*–*mphA* conferring resistance

to trimethoprim, aminoglycoside, quaternary ammonium compounds (disinfectant), sulphonamide, chromate (metal) and macrolides. Identical integrons have been previously observed in Australian clinical ST131 and ST200 isolates (39) as well as an ExPEC ST410 isolate from a canine in Portugal (82). Clade C2 isolates were also strongly associated with *bla*$_{\text{CTX-M-15}}$, *bla*$_{\text{OXA-1}}$ and the acquired aminoglycoside gene with extended fluroquinolone resistance *acc(6')–Ib–cr*, all of which are well documented in this clade globally (83–85). In clade C2, class one integrons, *bla*$_{\text{CTX-M-15}}$ and other AMR genes are most commonly carried on F2:A1:B− plasmids (34, 86, 87). Indeed, this pMLST is thought to have shaped the evolution of C2 (31). Interestingly, while F2:A1:B− was the most common pMLST in the Australian cohort of clade C2, it was only present in 20 isolates (19%). The remaining isolates carried F plasmids of 35 different pMLSTs, thus the previously reported strong correlation between C2 and F2:A1:B− plasmids was not observed in this ST131 population.

ST131 clade C1 is also associated with ESBL production worldwide and most commonly harbours *bla*$_{\text{CTX-M-14}}$ (17). However, a subpopulation of C1 isolates carrying *bla*$_{\text{CTX-M-27}}$ (C1-M27), providing greater ceftazidime resistance (88), was first identified in Japan in 2006 (60) and has since emerged globally and is being reported at increasing frequencies in both humans and animals (86, 89–92). In Australia, we found C1 isolates most frequently carried *bla*$_{\text{CTX-M-27}}$ (52%), while only 6% carried *bla*$_{\text{CTX-M-14}}$. Intriguingly, though typically reported as antibiotic sensitive (83), 71% of Australian clade A isolates carried at least one ESBL gene, of which *bla*$_{\text{CTX-M-27}}$ was most prevalent. While biased selection criteria cannot be ruled out as a cause of the high percentage of ESBL resistant clade A isolates observed here, we did note that several clade A isolates had chromosomal *bla*$_{\text{CTX-M-27}}$ genes while others shared the same genetic context as C1-M27 isolates. Furthermore, three clade A isolates possessed the dual *parC*-1aAB *gyrA*-1AB mutations previously only reported in clade C. These two mutations not only confer resistance to fluroquinolone but also improve fitness in the absence of antibiotics (93) and it is generally acknowledged that the combination of fluroquinolone and ESBL resistance propelled the global expansion of clade C (7). It is therefore alarming that clade A may be following suit and should be monitored closely, especially when population estimates have flagged clade A as the dominant clade in Australia (37), and unbiased sampling methodologies have also identified clade A as dominant in the community in PR China (94) and domestic wastewater in Canada (10).

In terms of virulence factors, ST131 isolates from Australia did not deviate from previous reports in that most (>90%) carried yersiniabactin receptor *fyuA*, uropathogen-specific protein *usp*, pathogenicity island marker *malx*, secreted autotransporter *sat*, aerobactin genes *iutA* and *iucD* and capsule *kpsM II* (6, 95). Notably the *kpsM II* K2 allele has previously been reported to be

strongly associated with clade A (6), however in this collection K5 was the most common allele in clade A and C1, K1 in C2 and the only two examples of K2 alleles originated from clade B.

Given that ColV and pUTI89-like plasmids are associated with ExPEC virulence (96) we performed plasmid mapping and ascertained that ST131 ColV plasmid carriage is low (*n*=10) in this Australian collection. ColV plasmids are a defining feature of APEC (19, 21) and the low carriage here may be due to the collection being largely human sourced, though the single poultry isolate did carry a ColV-like plasmid. Conversely, pUTI89-like cargo, which is linked to bacterial invasion (96), was present in approximately half of all clade A and B isolates, though only present in 3% of clade C isolates. The presence of pUTI89 has previously been reported to reduce the number of other plasmids in the ST95 ExPEC lineage and has not been reported as carrying ARGs (63). Here we found that the presence of pUTI89-like cargo was also associated with fewer replicons in ST131. Notably, we identified several isolates harbouring pUTI89-like plasmids that also carried $bla_{\text{CTX-M-15}}$ and $bla_{\text{CTX-M-27}}$, however, where genetic context could be ascertained, these ESBL genes were always located on the chromosome.

Lastly, a subclade of H27 ST131 isolates have been recently described as hybrid EAEC/ExPEC strains due to the acquisition of a pAA plasmid carrying genes that code for AAF which are central to EAEC pathogenesis and prolific biofilm formation (65, 97). A hybrid H27 sublineage has reportedly caused an outbreak of UTIs and bacteraemia in Denmark (65). Here we found that the only two H27 ST131 isolates in this collection (two human sourced clade B isolates) carried a pAA-like plasmid, though these isolates were not closely related to the Danish outbreak strains.

6.9.1. Limitations

In addition to the aforementioned skew towards human sourced ST131 isolates, this study also had the limitation of Enterobase genome selection by ST, Australian origin and year of isolation only. It is unknown whether these ST131 isolates had associated selection biases, such as ESBL screening or pathologies. Additionally, given that 43% of Enterobase-extracted isolates were isolated in 2017, no meaningful temporal studies could be conducted.

6.10. Concluding statement

In conclusion, our findings demonstrate that the ST131 population in Australia carry genes conferring resistances to multiple antibiotic classes, carry similar virulence cargo and indicate no niche associations. This highlights the importance of shifting emphasis from solely clinically focused sampling as it seems likely that strains can move across ecological niches. A better

understanding of ST131 and ExPEC reservoirs could lead to the development of preventative measures regarding the spread of infection and AMR.

6.11. References

1. Hastak P, Fourment M, Darling AE, Gottlieb T, Cheong E, et al. Escherichia coli ST8196 is a novel, locally evolved, and extensively drug resistant pathogenic lineage within the ST131 clonal complex. Emerg Microbes Infect 2020;9:1780–1792.
2. Pitout JDD, DeVinney R. Escherichia coli ST131: a multidrugresistant clone primed for global domination. F1000Res 2017;6:195.
3. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev 2014;27:543–574.
4. Banerjee R, Johnson JR. A new clone sweeps clean: the enigmatic emergence of Escherichia coli sequence type 131. Antimicrob Agents Chemother 2014;58:4997–5004.
5. Nicolas-Chanoine M-H, Blanco J, Leflon-Guibout V, Demarty R, Alonso MP, et al. Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. J Antimicrob Chemother 2008;61:273–281.
6. Dahbi G, Mora A, Mamani R, López C, Alonso MP, et al. Molecular epidemiology and virulence of Escherichia coli O16:H5-ST131: comparison with H30 and H30-Rx subclones of O25b:H4-ST131. Int J Med Microbiol 2014;304:1247–1257.
7. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, et al. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A 2014;111:5694–5699.
8. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, et al. The epidemic of extended-spectrum-β-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. mBio 2013;4:e00377-13.
9. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, et al. Evolutionary history of the global emergence of the Escherichia coli epidemic clone ST131. mBio 2016;7:e02162-15.
10. Finn TJ, Scriver L, Lam L, Duong M, Peirano G, et al. A comprehensive account of Escherichia coli sequence type 131 in wastewater reveals an abundance of fluoroquinolone-resistant clade A strains. Appl Environ Microbiol 2020;86:e01913-19.
11. Nickel JC. Practical management of recurrent urinary tract infections in premenopausal women. Rev Urol 2005;7:11–17.
12. Raz R, Chazan B, Kennes Y, Colodner R, Rottensterich E, et al. Empiric use of trimethoprim-sulfamethoxazole (TMP-SMX) in the treatment of women with uncomplicated urinary tract infections, in a geographical area with a high prevalence of TMP-SMXresistant uropathogens. Clin Infect Dis 2002;34:1165–1169.
13. Le Saux N, Robinson J. Aminoglycosides—alive and well in treatment of pediatric infections: a case of benefit versus risk. Official J Assoc Med Microbiol Infect Dis Canada 2019;4:1–5.
14. Kudinha T, Johnson JR, Andrew SD, Kong F, Anderson P, et al. Escherichia coli sequence type 131 as a prominent cause of antibiotic resistance among urinary Escherichia coli isolates from reproductive-age women. J Clin Microbiol 2013;51:3270–3276.
15. Ludden C, Decano AG, Jamrozy D, Pickard D, Morris D, et al. Genomic surveillance of Escherichia coli ST131 identifies local expansion and serial replacement of subclones. Microb Genom 2020;6.

16. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of Escherichia coli ST131. mBio 2016;7:e00347-16.
17. Decano AG, Downing T. An Escherichia coli ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. Sci Rep 2019;9:17394.
18. Reid CJ, McKinnon J, Djordjevic SP. Clonal ST131-H22 Escherichia coli strains from a healthy pig and a human urinary tract infection carry highly similar resistance and virulence plasmids. Microb Genom 2019;5.
19. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, et al. Escherichia coli ST131-H22 as a foodborne uropathogen. mBio 2018;9:e00470-18.
20. Roer L, Overballe-Petersen S, Hansen F, Johannesen TB, Stegger M, et al. ST131 fimH22 Escherichia coli isolate with a blaCMY-2 /IncI1/ST12 plasmid obtained from a patient with bloodstream infection: highly similar to E. coli isolates of broiler origin. J Antimicrob Chemother 2019;74:557–560.
21. Cummins ML, Reid CJ, Roy Chowdhury P, Bushell RN, Esbert N, et al. Whole genome sequence analysis of Australian avian pathogenic Escherichia coli that carry the class 1 integrase gene. Microb Genom 2019;5.
22. Johnson TJ, Siek KE, Johnson SJ, Nolan LK. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian Escherichia coli strains. J Bacteriol 2006;188:745–758.
23. Tivendale KA, Logue CM, Kariyawasam S, Jordan D, Hussein A, et al. Avian-pathogenic Escherichia coli strains are similar to neonatal meningitis E. coli strains and are able to cause meningitis in the rat model of human disease. Infect Immun 2010;78:3412–3419.
24. Skyberg JA, Johnson TJ, Johnson JR, Clabots C, Logue CM, et al. Acquisition of avian pathogenic Escherichia coli plasmids by a commensal E. coli isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. Infect Immun 2006;74:6287–6292.
25. 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 (Reading) 2009;155:450–460.
26. McKinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant Escherichia coli ST58 causing urosepsis. Int J Antimicrob Agents 2018;52:430–435.
27. McKinnon J, Roy Chowdhury P, Djordjevic SP. Molecular analysis of an IncF ColV-like plasmid lineage that carries a complex resistance locus with a trackable genetic signature. Microb Drug Resist 2020;26:787–793.
28. Cointe A, Birgy A, Mariani-Kurkdjian P, Liguori S, Courroux C, et al. Emerging multidrug-resistant hybrid pathotype shiga toxinproducing Escherichia coli o80 and related strains of clonal complex 165, Europe. Emerg Infect Dis 2018;24:2262–2269.
29. Saidenberg ABS, Stegger M, Price LB, Johannesen TB, Aziz M, et al. mcr-positive Escherichia coli ST131-H22 from poultry in Brazil. Emerg Infect Dis 2020;26:1951–1954.
30. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, et al. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res 2017;27:1437–1449.
31. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, et al. Separate F-Type plasmids have shaped the evolution of the H30 subclone of Escherichia coli sequence Type 131. mSphere 2016;1:e00121-16.
32. Mahon BM, Brehony C, Cahill N, McGrath E, O'Connor L, et al. Detection of OXA-48-like-producing Enterobacterales in Irish recreational water. Sci Total Environ 2019;690:1–6.

33. Peirano G, Schreckenberger PC, Pitout JDD. Characteristics of NDM-1-producing Escherichia coli isolates that belong to the successful and virulent clone ST131. Antimicrob Agents Chemother 2011;55:2986–2988.
34. Mahérault A-C, Kemble H, Magnan M, Gachet B, Roche D, et al. Advantage of the F2:A1:B- IncF pandemic plasmid over IncC plasmids in in vitro acquisition and evolution of blaCTX-M genebearing plasmids in Escherichia coli. Antimicrob Agents Chemother 2019;63:10.
35. Hayashi M, Matsui M, Sekizuka T, Shima A, Segawa T, et al. Dissemination of IncF group F1:A2:B20 plasmid-harbouring multidrugresistant Escherichia coli ST131 before the acquisition of blaCTX-M in Japan. J Glob Antimicrob Resist 2020;23:456–465.
36. Villa L, García-Fernández A, Fortini D, Carattoli A. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. J Antimicrob Chemother 2010;65:2518–2529.
37. Rogers BA, Ingram PR, Runnegar N, Pitman MC, Freeman JT, et al. Sequence type 131 fimH30 and fimH41 subclones amongst Escherichia coli isolates in Australia and New Zealand. Int J Antimicrob Agents 2015;45:351–358.
38. Rogers BA, Ingram PR, Runnegar N, Pitman MC, Freeman JT, et al. Community-onset Escherichia coli infection resistant to expandedspectrum cephalosporins in low-prevalence countries. Antimicrob Agents Chemother 2014;58:2126–2134.
39. Li D, Reid CJ, Kudinha T, Jarocki VM, Djordjevic SP. Genomic analysis of trimethoprim-resistant extraintestinal pathogenic Escherichia coli and recurrent urinary tract infections. Microb Genom 2020;6.
40. Hastak P, Cummins ML, Gottlieb T, Cheong E, Merlino J, et al. Genomic profiling of Escherichia coli isolates from bacteraemia patients: a 3-year cohort study of isolates collected at a Sydney teaching hospital. Microb Genom 2020;6.
41. Kidsley AK, White RT, Beatson SA, Saputra S, Schembri MA, et al. Companion animals are spillover hosts of the multidrug-resistant human extraintestinal Escherichia coli pandemic clones ST131 and ST1193. Front Microbiol 2020;11:1968.
42. Mukerji S, Stegger M, Truswell AV, Laird T, Jordan D, et al. Resistance to critically important antimicrobials in Australian silver gulls (Chroicocephalus novaehollandiae) and evidence of anthropogenic origins. J Antimicrob Chemother 2019;74:2566–2574.
43. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 2015;43:e15.
44. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 2010;5:e9490.
45. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 2019;47:W256–W259.
46. Yu G, Smith DK, Zhu H, Guan Y, Lam TT, et al. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 2016;8:28–36.
47. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 2020;37:1530–1534.
48. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3693.
49. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res 2019;47:5539–5549.
50. Roer L, Tchesnokova V, Allesøe R, Muradova M, Chattopadhyay S, et al. Development of a web tool for Escherichia coli subtyping based on fimH alleles. J Clin Microbiol 2017;55:2538–2543.

51. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 2020;75:3491–3500.
52. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, et al. In silico detection and typing of plasmids using plasmid finder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 2014;58:3895–3903.
53. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–2849.
54. Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using the CGView family of tools. Brief Bioinform 2019;20:1576–1582.
55. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–2069.
56. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 2016;17:238.
57. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, et al. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics 2018;34:292–293.
58. Blanco J, Mora A, Mamani R, López C, Blanco M, et al. Four main virotypes among extended-spectrum-β-lactamase-producing isolates of Escherichia coli O25b:H4-B2-ST131: bacterial, epidemiological, and clinical characteristics. J Clin Microbiol 2013;51:3358–3367.
59. Manges AR. Escherichia coli causing bloodstream and other extraintestinal infections: tracking the next pandemic. Lancet Infect Dis 2019;19:1269–1270.
60. Matsumura Y, Pitout JDD, Gomi R, Matsuda T, Noguchi T, et al. Global Escherichia coli sequence type 131 clade with blaCTX-M-27 Gene. Emerg Infect Dis 2016;22:1900–1907.
61. Sáenz Y, Zarazaga M, Briñas L, Ruiz-Larrea F, Torres C. Mutations in gyrA and parC genes in nalidixic acid-resistant Escherichia coli strains from food products, humans and animals. J Antimicrob Chemother 2003;51:1001–1005.
62. Gillings MR. Integrons: past, present, and future. Microbiol Mol Biol Rev 2014;78:257–277.
63. 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:e00390-16.
64. Kaur P, Chakraborti A, Asea A. Enteroaggregative Escherichia coli: an emerging enteric food borne pathogen. Interdiscip Perspect Infect Dis 2010;2010:254159.
65. Boll EJ, Overballe-Petersen S, Hasman H, Roer L, Ng K, et al. Emergence of enteroaggregative Escherichia coli within the ST131 lineage as a cause of extraintestinal infections. mBio 2020;11:e00353-20.
66. Sidjabat HE, Derrington P, Nimmo GR, Paterson DL. Escherichia coli ST131 producing CTX-M-15 in Australia. J Antimicrob Chemother 2010;65:1301–1303.
67. Flament-Simon S-C, de Toro M, Mora A, García V, García-Meniño I, et al. Whole genome sequencing and characteristics of mcr-1-harboring plasmids of porcine Escherichia coli isolates belonging to the high-risk clone O25b:H4-ST131 Clade B. Front Microbiol 2020;11:387.
68. Jamborova I, Johnston BD, Papousek I, Kachlikova K, Micenkova L, et al. Extensive genetic commonality among wildlife, wastewater, community, and nosocomial isolates of Escherichia coli sequence type 131 (H30R1 and H30Rx Subclones) that Carry blaCTX-M-27 or blaCTX-M-15. Antimicrob Agents Chemother 2018;62:e00519-18.
69. McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet 2016;12:e1006280.

70. Mills EG, Martin MJ, Luo TL, Ong AC, Maybank R, Corey BW, Harless C, Preston LN, Rosado-Mendez JA, Preston SB, Kwak YI, Backlund MG, Bennett JW, Mc Gann PT, Lebreton F. A one-year genomic investigation of Escherichia coli epidemiology and nosocomial spread at a large US healthcare network. Genome Med. 2022 Dec 30;14(1):147.
71. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad Sci U S A. 2012 Oct 9;109(41):E2774-83.
72. Reeves PR, Liu B, Zhou Z, Li D, Guo D, Ren Y, Clabots C, Lan R, Johnson JR, Wang L. Rates of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS One. 2011;6(10):e26907.
73. Ingle DJ, Howden BP, Duchene S. Development of phylodynamic methods for bacterial pathogens phylodynamic methods for bacterial pathogens. Trends Microbiol 2021;29:797.
74. Rodríguez-Beltrán J, Tourret J, Tenaillon O, López E, Bourdelier E, et al. High recombinant frequency in extraintestinal pathogenic Escherichia coli strains. Mol Biol Evol 2015;32:1708–1716.
75. Paul S, Linardopoulou EV, Billig M, Tchesnokova V, Price LB, et al. Role of homologous recombination in adaptive diversification of extraintestinal Escherichia coli. J Bacteriol 2013;195:231–242.
76. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. Nat Rev Genet 2018;19:549–565.
77. Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS. Genetic diversity and population structure of Escherichia coli isolated from freshwater beaches. Environ Microbiol. 2007 Sep;9(9):2274-88.
78. Salerno A, Delétoile A, Lefevre M, Ciznar I, Krovacek K, Grimont P, Brisse S. Recombining population structure of Plesiomonas shigelloides (Enterobacteriaceae) revealed by multilocus sequence typing. J Bacteriol. 2007 Nov;189(21):7808-18.
79. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 2009 Feb;3(2):199-208.
80. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. J Antimicrob Chemother 2017;72:2145–2155.
81. Forde BM, Roberts LW, Phan M-D, Peters KM, Fleming BA, et al. Population dynamics of an Escherichia coli ST131 lineage during recurrent urinary tract infection. Nat Commun 2019;10:3643.
82. Brilhante M, Menezes J, Belas A, Feudi C, Schwarz S, et al. OXA-181-producing extraintestinal pathogenic Escherichia colisequence type 410 isolated from a dog in Portugal. Antimicrob Agents Chemother 2020;64.
83. Pitout JDD, Finn TJ. The evolutionary puzzle of Escherichia coli ST131. Infect Genet Evol 2020;81:104265.
84. Alsharapy SA, Yanat B, Lopez-Cerero L, Nasher SS, Díaz-De-Alba P, et al. Prevalence of ST131 clone producing both ESBL CTX-M-15 and AAC(6')Ib-cr among ciprofloxacin-resistant Escherichia coli isolates from Yemen. Microb Drug Resist 2018;24:1537–1542.
85. Livermore DM, Day M, Cleary P, Hopkins KL, Toleman MA, et al. OXA-1 β-lactamase and non-susceptibility to penicillin/βlactamase inhibitor combinations among ESBL-producing Escherichia coli. J Antimicrob Chemother 2019;74:326–333.
86. Blanc V, Leflon-Guibout V, Blanco J, Haenni M, Madec J-Y, et al. Prevalence of day-care centre children (France) with faecal CTX-Mproducing Escherichia coli comprising O25b:H4 and O16:H5 ST131 strains. J Antimicrob Chemother 2014;69:1231–1237.

87. Zhang L, Lü X, Zong Z. The emergence of blaCTX-M-15-carrying Escherichia coli of ST131 and new sequence types in Western China. Ann Clin Microbiol Antimicrob 2013;12:35.
88. Bonnet R, Recule C, Baraduc R, Chanal C, Sirot D, et al. Effect of D240G substitution in a novel ESBL CTX-M-27. J Antimicrob Chemother 2003;52:29–35.
89. Birgy A, Levy C, Nicolas-Chanoine M-H, Cointe A, Hobson CA, et al. Independent host factors and bacterial genetic determinants of the emergence and dominance of Escherichia coli sequence type 131 CTX-M-27 in a community pediatric cohort study. Antimicrob Agents Chemother 2019;63.
90. Ghosh H, Doijad S, Falgenhauer L, Fritzenwanker M, Imirzalioglu C, et al. blaCTX-M-27 – Encoding Escherichia coli sequence type 131 lineage C1-M27 clone in clinical isolates, Germany. Emerg Infect Dis 2017;23:1754–1756.
91. Melo LC, Haenni M, Saras E, Duprilot M, Nicolas-Chanoine M-H, et al. Emergence of the C1-M27 cluster in ST131 Escherichia coli from companion animals in France. J Antimicrob Chemother 2019;74:3111–3113.
92. Zendri F, Maciuca IE, Moon S, Jones PH, Wattret A, et al. Occurrence of ESBL-producing Escherichia coli ST131, including the H30-Rx and C1-M27 subclones, among urban seagulls from the United Kingdom. Microb Drug Resist 2020;26:697–708.
93. Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. Longterm experimental evolution in Escherichia coli. XII. DNA topology as a key target of selection. Genetics 2005;169:523–532.
94. Zhong Y-M, Liu W-E, Liang X-H, Li Y-M, Jian Z-J, et al. Emergence and spread of O16-ST131 and O25b-ST131 clones among faecal CTX-M-producing Escherichia coli in healthy individuals in Hunan Province, China. J Antimicrob Chemother 2015;70:2223–2227.
95. Downing T. Tackling drug resistant infection outbreaks of global pandemic Escherichia coli ST131 using evolutionary and epidemiological genomics. Microorganisms 2015;3:236–267.
96. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect Immun 2010;78:1457–1467.
97. Boll EJ, Ayala-Lujan J, Szabady RL, Louissaint C, Smith RZ, et al. Enteroaggregative Escherichia coli adherence fimbriae drive inflammatory cell recruitment via interactions with epithelial MUC1. mBio 2017;8:e00717-17.

Chapter 7: General Discussion and Future Directions

Uropathogenic *E. coli* are costly, reduce patient quality of life and can lead to the development of more life-threatening conditions like urosepsis. Additionally, UTIs are one of the most common bacterial infections and are often treated with antibiotics which enhances AMR dissemination and evolution. It is vital to know the genetic characteristics of bacteria causing UTIs to understand how AMR is changing and to ascertain change, retrospective studies are required, especially before the advent of Illumina MiSeq NGS in 2011, which made sequencing far more affordable (1). Most of the isolates used in this retrospective study were from 2006 (81.6%), when WGS would have cost an ~USD$3150 per *E. coli* genome (2). AMR has been rising in Australia in the last few years (3) and the landscape of lineages present and dominating in the UPEC population is constantly changing. A recent 12-year study of *E. coli* population dynamics from French BSI patients highlights, that on the phylogroup level populations remains stable but ST level showed noticeable changes, and some STs exhibits highly dynamic population variation sometimes coupled with virulence gene acquisition (4). Therefore, it is important to have a baseline for clinical population of UPEC to be able to study how AMR evolves.

Equally important is to identify and study local transmission events like outbreaks to provide better policies for prevention of patient cross contamination. This research could also provide insight into the genomic bases of uropathogenicity, and possibly future ways to counteract it. Most modern genomic epidemiological studies of UPEC focus on antibiotic-resistant lineages, such as ST131, due to their enhanced clinical significance. However, this leaves STs and lineages which are not resistant or not MDR poorly understood, obfuscating the disease burden of these lineages. This limitation also reduces the capacity for identification of emerging pathogens and limiting our understanding of the genomic context from which pandemic ExPEC lineages emerge.

Given that each results chapter includes a discussion section, this final chapter offers a holistic analysis of the thesis goals, along with a critical evaluation of the study limitations and future research directions.

7.1. Aim 1

Determine the population structure and MLSTs of a retrospective collection of urine-sourced *E. coli* isolates from an Australian rural hospital.

This aim was achieved in Chapters 4 and 5. Chapter 4 provided the population structure of a non-antibiotic selected collection (n=336), while chapter 5 covered the populational structure of a trimethoprim resistant collection (n=67) from same hospital and around same time frame.

In both collections the primary patient diagnosis was cystitis, followed by pyelonephritis and urosepsis. Our antibiotic selection free collection found a high prevalence of phylogroup B2 isolates with the top STs being ST73 > ST95 > ST127 > ST131, which accounted for 43% of all isolates. Conversely, in the trimethoprim resistant the top STs were ST131 > ST69 > ST38 ≥ ST95 (Figure 7.1). The top STs in the trimethoprim resistant collection were more similar to the top pandemic ExPEC STs flagged in a comprehensive systematic review of ExPEC (5), that is ST131 > ST69 > ST10 > ST405 > ST38, and the paper acknowledged that most of the included studies were based on AMR phenotypes. This focus on antibiotic-selected collections skews actual ExPEC populational structures and likely misappropriates extraintestinal disease burden to lineages associated with MDR and ESBL resistance like ST131, ST405 and ST38 (5). In addition, an AMR bias limits a one health understanding of ExPEC disease and clouds understanding of ExPEC reservoirs.

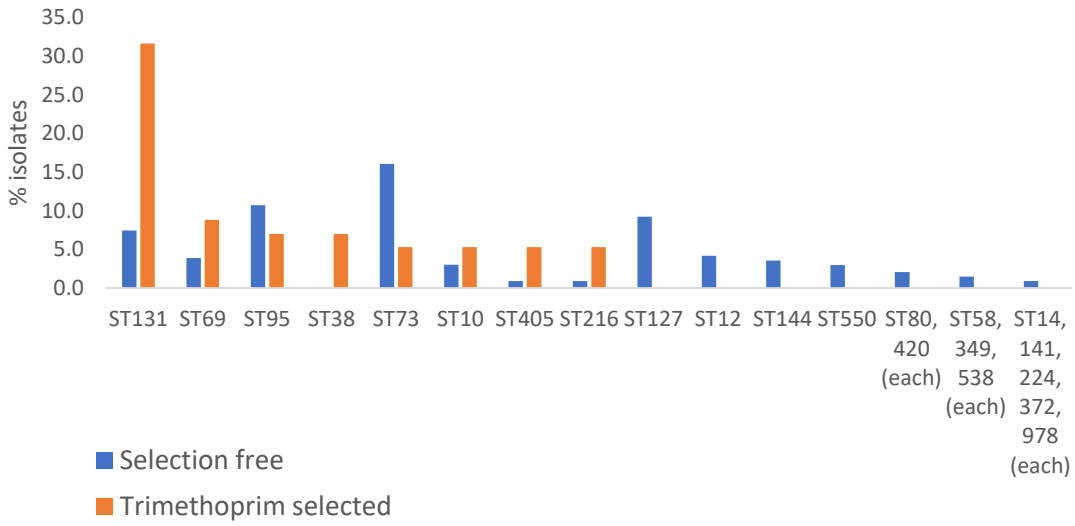


Figure 7.1: Distribution of STs withing selection free and trimethoprim selected collections. Only STs with more than three isolates are present here, some STs with equal number of isolates are represented as one bar. Isolates from serial patients studied in Chapter 5 that were not phenotypically resistant to trimethoprim were removed from this graph.

In the AMR-selection free cohort, ST73 was the most prevalent (16% vs 5% in trimethoprim resistant collection). ST73 belongs to phylogroup B2 (one commonly associated with human ExPEC STs) and is a documented cause of clinical UTI and sepsis in humans in Europe, Australia and US (6-8). Similar to our results, studies from the US and Europe report that ST73 is the most common ExPEC ST isolated from such pathologies in the absence of antibiotic resistance selection criteria; however, this ST was not commonly MDR or resistant to critically important antibiotics, thus leading to less interest from scientific community compared to lineages like ST131. In this study, ST73 isolates did indeed have low carriage of ESBL genes (2%), but genes

conferring resistances to aminoglycosides, sulphonamides and NSBL were much more abundant. With the recent introduction of the One Health WHO Tricycle Surveillance protocol (9), which implements the monitoring of a single indicator organism – ESBL-producing *E. coli* – it is likely that important UPEC lineages such as ST73 will continue to be underreported. From a clinical standpoint, it might be argued that lineages such as ST73 can be readily treated with extended-spectrum beta-lactams and therefore take a backseat to treatment resistant lineages. But that could be considered a little short sighted, given the propensity of *E. coli* to acquire ARGs through HGT. Lest we forget that ST131 was not always ESBL-producing.

In addition to ST73, ST95 and ST127 may continue to be underreported due to relatively low AMR, despite being highly virulent. ST95 is known to be one of the dominant ExPEC lineages and often causes UTIs and BSIs in humans and colibacillosis in poultry. ST95 was found to be the second most prevalent and covered 11% of the selection free collection (vs 6% in trimethoprim resistant collection), which is similar to other studies (10, 11). ST95 is known to have high zoonotic potential and is postulated as being associated with food born UTIs (12); however, in comparison to some ST73, ST127 and ST131 isolates, our phylogenomic and SNP analyses found that OBH ST95 isolates tended to cluster with other human ST95 isolates and only one OBH ST95 was closely related (< 100 SNPs) to an isolate from a non-human source (poultry). ST127 is an emerging uropathogen frequently isolated from humans as well as companion animals (13, 14). We found six OBH isolates (19%) that had < 100 SNPs to isolates obtained from companion animals, adding weight to the hypothesis that companion animals, particularly dogs, may be a reservoir for ST127. ST73 and ST127 are known to reside in companion animals and evidence of interspecies transmission in Australia has been reported (14, 15).

Many UPEC WGS studies are concentrated on *E. coli* ST131 (detailed in Chapter 6), due to its high rates of MDR and ESBL resistance. *E. coli* ST131 is a prominent and globally disseminated ExPEC which is a common cause of antibiotic resistant hospital and community-acquired UTIs and BSIs. While an important ST in and of itself, researchers should not neglect non ESBL resistant pandemic lineages (a literature search of PubMed Central search at 01/11/22; "*E. coli* ST73" returns 400 results while "*E. coli* ST131" returns 3308 results).

The population structure and MLST analysis conducted in Aim 1 provided valuable insights into the distribution of STs within the urine-sourced *E. coli* isolates. This information laid the foundation for further investigation in Aim 2, where WGS data and genotyping tools were utilised to identify ARGs, VAGs, plasmid RSTs, and critical MGEs within the isolates.

7.2. Aim 2

Identify ARGs, VAGs, plasmid RSTs and critical MGEs using WGS data, multiple databases, read mapping and genotyping tools.

This aim was addressed in detail in every result chapter. A brief overview of some major findings include:

1) An overall low level of ARG carriage for antibiotic selection free collection detected.
2) Two or more times higher overall ARG carriage in trimethoprim resistance collection and ST131 in particular.
3) Australian *E. coli* ST131 clade A exhibiting high carriage of AMR and ESBL genes and examples of the dual *parC*-1aAB *gyrA*-1AB mutations corresponding to fluoroquinolone resistance, previously only reported in clade C.
4) Though now dominant in Australian ST131 C1 and A clades (chapter 6), $bla_{\text{CTX-M-27}}$ genes were not detected in any of the 66 ST131 isolates from OBH, indicating that this gene had not yet reached this population before 2011.
5) High carriage of integrons associated with MDR.
6) High carriage of ubiquitous ExPEC-associated VAGs, including those encoding iron acquisition systems, adhesins, invasins, and toxins.

While reporting on ARGs and VAGs are the "bread and butter" of genomic studies, this body of work did attempt to link ARGs with plasmid carriage. Previous studies have highlighted F plasmids that carry *cjrABC-senB* and ColV virulence plasmids, which carry an impressive arsenal of iron-acquisition operons, toxins, and autotransporter genes. These two F plasmid lineages dominate ExPEC populations globally. F virulence plasmids carrying *senB-cjrABC* virulence genes were a feature of our selection free collection. These *senB-cjrABC*+ plasmids were split into two categories: pUTI89-like and non-pUTI89-like. Compared to all other plasmid replicons, isolates with pUTI89-like plasmids carried fewer ARGs, whilst isolates with *senB-cjrABC*+/non-pUTI89 plasmids had a significantly higher load of ARGs and class 1 integrons.

ColV plasmids are known to play important roles in development of BSIs (16), due to their iron acquisition systems (4), however in both of collections ColV plasmids were not heavily represented. This could be attributed to the fact that, even though 4.2% of isolates were obtained from patients diagnosed with sepsis, the samples were urine, not blood, so the isolates may not have been the aetiological agents.

As these plasmid association findings are discussed in corresponding chapters, here I will mostly focus on related limitations. First is the limitation of short read sequencing. Plasmid and integron studies in all three result chapters are limited by this sequencing method as it only provides contigs and it is rarely possible to establish full plasmid or transposon structures. Some isolates found carrying the same integrons carried different plasmid replicons leading to the assumption that integrons or transposons containing them have transferred between different plasmids or to the chromosome. Even with this limitation in place, it was possible to perform a limited study of integrons as vectors for some ARGs, PAIs and plasmid types. But this research should be considered with a degree of caution as no long-read sequencing was done to corroborate these findings. Plasmid and PAI mapping results presented in this thesis have the limitation of only confirming gene presence, but not necessarily the same gene configuration or context as the reference used.

The comprehensive analysis of ARGs, VAGs, and MGEs performed in Aim 2 allowed for a direct comparison between isolates selected for trimethoprim resistance and those not selected for phenotypic antibiotic resistance. This comparison, as addressed in Aim 3, provided insights into the differences in populational structure, ARGs, VAGs, and MGEs between these two collections.

7.3. Aim 3

Compare isolates selected for trimethoprim resistance to those not selected for phenotypic antibiotic resistance, in terms of populational structure, ARGs, VAGs, and MGEs.

The populational structure comparison of these two collections was discussed in section 7.1, though the limitation of such a comparison should be mentioned, namely related to sample sizes - the trimethoprim selected collection is only a fifth of the size of the selection free collection which does not provide enough power for a thorough statistical comparison, hence the two collections were presented as two separate manuscripts. These limitations extend to ARG, VAG and MGEs (here, class 1 integrons) comparisons between the two collections. Nevertheless, some interesting initial observations can be made for the purpose of this thesis discussion.

Regarding ARGs, unsurprisingly, the collection selected for trimethoprim resistance exhibited high carriage of genes conferring resistance to trimethoprim. However, we also observed that this group carried more ARGs outside of those associated with resistance to trimethoprim (Figure 7.2). This could be attributed to high prevalence of ST131 lineage (Figure 7.1), well known to be associated with MDR (17), and also to higher class 1 integron carriage (Figure 7.2), which was to be expected since *dfrA* genes, the most common mechanism for trimethoprim resistance, are almost exclusively found within class 1 integrons (18). However, we did observe that 17% of

trimethoprim resistant isolates did not carry *dfrA* genes and 27% did not carry *intI1,* hinting at a yet undiscovered mechanism behind trimethoprim resistance.

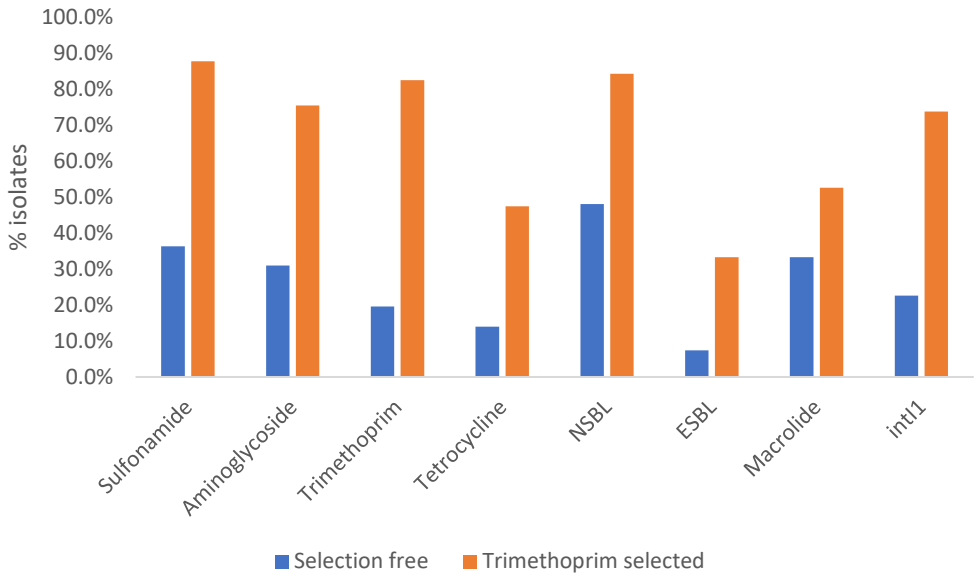


Figure 7.2: Genotypic AMR profiles for chapter 4 and 5 collections. Compares the percentage of isolates carrying AMR genes or *intI1* between the selection free collection and the trimethoprim resistant collection. Isolates from patients which were sampled multiple times, studied in Chapter 5, were removed from this graph.

The two collections also differed in terms of the most common class 1 integron structures. In the trimethoprim resistant isolates *intI1-dfrA12-aadA2-qacEΔ1-sul1*-ORF-*chrA-padR*-IS*6100-mphR-mrx-mphA* was more common, present in 17.9% of isolates. This structure is heavily associated with ST131s (chapter 6) and has recently been identified in ETEC (19). In the collection without AMR selection, the most prevalent was IS*26*-ORF-*intI1-aadA1-qacEΔ1-sul1*-ORF-IS*1326*-IS*1353*-Tn*21*-ORF-*merE-merD-merAΔ-bla*$_{TEM1}$-Tn*2-merR.* This integron and associated CRR is similar to Tn*4* (though Tn*4* does not possess IS*1353*). Tn*4* is an evolved version of the widely disseminated Tn*21,* with an insertion carrying *bla*$_{TEM-1}$ Tn3 in the *mer* operon (20).

VAGs profiles across both collections were quite similar, with common ExPEC genes present in both (Figure 7.3). The most prevalent genes encoded iron acquisition systems. As for many other lifeforms, iron is critical nutrient for growth of bacteria. Siderophores have a high affinity for iron, allowing bacterial cells to outcompete host cells in iron uptake. Certain siderophore genes, such as *eitA*, *fecA* and *iroN*, were more prominent in the trimethoprim selected collection (~10% higher carriage), possibly due to selective pressure exerted by the frequent use of trimethoprim in treating UTIs caused by *E. coli*. Strains carrying iron-acquisition genes have a competitive advantage in iron-limited environments like the urinary tract. This higher prevalence of

siderophore genes could also be attributed to the presence of plasmids, which often carry both resistance and virulence genes, providing an additional mechanism for their persistence and spread in the population. Further research is needed to fully elucidate the relationship between antibiotic selection, plasmids, and the prominence of siderophore genes in UTI-sourced *E. coli*.

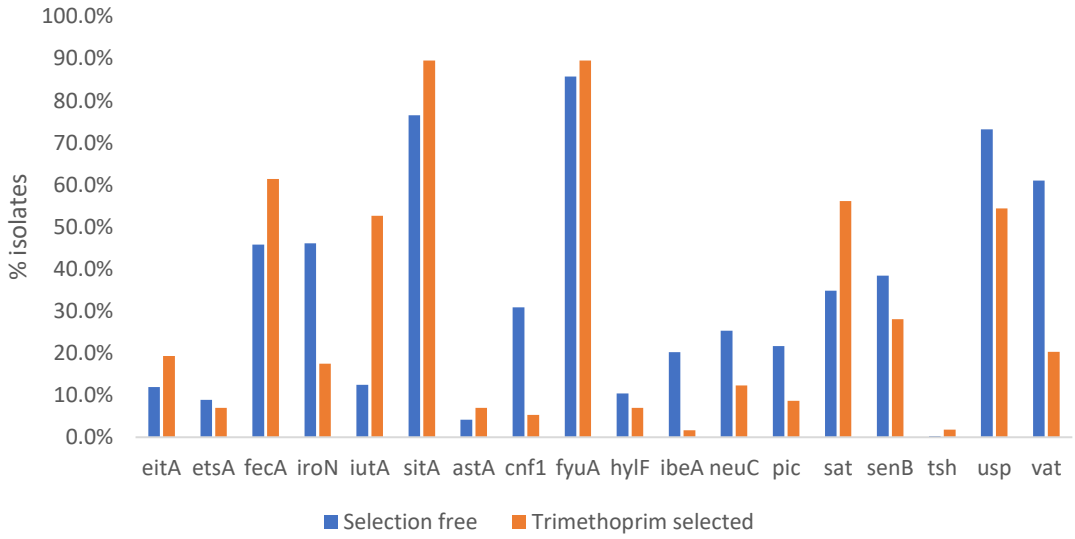


Figure 7.3 Prevalence of selected VAGs across two collections. Isolates from patients which were sampled multiple times, studied in Chapter 5, were removed from this graph.

In Aim 3, the comparison of isolates selected for trimethoprim resistance and those not selected for antibiotic resistance revealed interesting observations regarding populational structure, ARGs, VAGs, and MGEs. Building upon these findings, Aim 4 aimed to determine any statistically significant relationships between host age, sex, or clinical syndrome and the profiles of ARGs and VAGs.

7.4. Aim 4

Determine any statistically significant relationships between host age, sex, or clinical syndrome regarding ARG and VAG profiles.

This aim was addressed in Chapter 4. Multiple statistical analyses were performed to identify possible relationships between host factors and ARG and VAG profile (in terms of the types of such genes present or the count of such genes present), however no statistically significant relationships could be identified between host factors and ARGs and VAGs vs sex or uroclinical syndrome. However, isolates obtained from children had significantly more VAGs than isolates sourced from adults, providing an interesting point for future research.

Limitations related to this aim include that this analysis was performed on samples from only one hospital servicing a rural and regional area of Australia. This fact highly limits our ability to

generalise these trends to the majority of the population of Australia as they live in an urban environment where they do not have access to poultry and cattle farms which serve as a pool for AMR and MDR isolates (21, 22). Additionally, there were limitations in metadata, as no information regarding patient co-morbidities, treatments and outcomes were available, making it impossible to establish direct links to genes presence/absence and disease severity or possible mortality. Without information about the treatments administered, it becomes challenging to assess the significance of the ARG profiles.

7.5. Future perspectives and directions

Potential future work could involve performing WGS on isolates sourced from the same hospital but from a later date. This will allow us to conduct temporal studies and see how the landscape of ExPEC has changed at least within the area of just one rural Australian hospital. In combination, long-read sequencing on selected isolates of interest could be used to further study AMR-associated plasmids and other MGEs.

Further research could also involve more participating hospitals especially from high population density areas of the country, to provide comparisons between a rural and metropolitan population of *E. coli*. Ideally, the establishment of a national microbial genomic surveillance system would be beneficial in enhancing our understanding of ExPEC in Australia and overseas. Efforts in establishing such a surveillance system for ExPEC and other bacterial pathogens are currently underway, such as the AusTrakka initiative (cdgn.org.au/austrakka), Australia's nationally endorsed genomic surveillance platform (23).

Future research should conduct longitudinal studies to track the changes in the genetic characteristics of UPEC over time would providing valuable insights into the evolution of AMR and virulence. Additionally, adopting a One Health approach that considers the interactions between humans, animals, and the environment would enhance our understanding of the transmission dynamics of UPEC. Investigating potential reservoirs, such as companion animals, and studying transmission events, like outbreaks, will contribute to the development of preventive strategies and policies to reduce patient cross-contamination.

Public health agencies should review their policies regarding the use of antibiotics, particularly in food production and companion animals. In addition, there is a need to perform routine genomic AMR profile testing on a larger number of E. coli isolates. While it is acknowledged that WGS of all *E. coli* isolates is impractical for diagnostic laboratories, increased availability of such data would greatly enhance the ability to predict potential evolutionary shifts in ExPEC.

Finally, it should be noted that this thesis provides extensive data on ARG carriage in *E. coli* from Australia in 2006, which could be later used to perform a proper comparison to *E. coli* isolates from other countries of similar isolation time. While obtaining 16-year-old *E. coli* isolates may present its own challenges, it is necessary to study these retrospective collections as they may help illuminate the evolutionary pathways of modern *E. coli* populations and how they have adapted to past human interventions. This in turn will assist in producing policies which can prevent the escalation of AMR in bacteria.

In conclusion, this study highlights the concern of ARG in ExPEC isolates. The presence of class 1 and 2 integrons, along with their association with ARG, further emphasizes the potential for multidrug resistance dissemination. Additionally, the identification of numerous VAGs underscores the evolving nature of ExPEC strains and their ability to cause UTIs. These findings emphasize the urgent need for comprehensive surveillance, prudent use of antimicrobials, and effective infection control measures to mitigate the emergence and spread of multidrug-resistant ExPEC strains, thereby safeguarding public health.

7.6. References

1. Balloux F, Brønstad Brynildsrud O, van Dorp L, Shaw LP, Chen H, Harris KA, et al. From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic. Trends Microbiol. 2018;26(12):1035-48. doi: 10.1016/j.tim.2018.08.004.
2. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [02/11/22]. Available from: www.genome.gov/sequencingcostsdata.
3. AURA 2021: Fourth Australian report on antimicrobial use and resistance in human health: Australian Commission on Safety and Quality in Health Care (ACSQHC); 2021 [Available from: https://www.safetyandquality.gov.au/sites/default/files/2021-09/aura_2021_-_report_-_final_accessible_pdf_-_for_web_publication.pdf.
4. Royer G, Darty MM, Clermont O, Condamine B, Laouenan C, Decousser JW, et al. Phylogroup stability contrasts with high within sequence type complex dynamics of Escherichia coli bloodstream infection isolates over a 12-year period. Genome Med. 2021;13(1):77. doi: 10.1186/s13073-021-00892-0.
5. 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). doi: 10.1128/cmr.00135-18.
6. Alhashash F, Wang X, Paszkiewicz K, Diggle M, Zong Z, McNally A. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR Escherichia coli ST73: high levels of genomic and plasmid diversity in causative isolates. Journal of Antimicrobial Chemotherapy. 2015;71(2):339-43. doi: 10.1093/jac/dkv365.
7. Adams-Sapper S, Diep BA, Perdreau-Remington F, Riley LW. Clonal composition and community clustering of drug-susceptible and -resistant Escherichia coli isolates from bloodstream infections. Antimicrob Agents Chemother. 2013;57(1):490-7. doi: 10.1128/aac.01025-12.
8. 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). doi: 10.1099/mgen.0.000255.

9. WHO integrated global surveillance on ESBL-producing E. coli using a “One Health” approach: implementation and opportunities: World Health Organization; 2021.
10. Matsui Y, Hu Y, Rubin J, de Assis RS, Suh J, Riley LW. Multilocus sequence typing of Escherichia coli isolates from urinary tract infection patients and from fecal samples of healthy subjects in a college community. Microbiologyopen. 2020;9(6):1225-33. doi: 10.1002/mbo3.1032.
11. Day MJ, Doumith M, Abernethy J, Hope R, Reynolds R, Wain J, et al. Population structure of Escherichia coli causing bacteraemia in the UK and Ireland between 2001 and 2010. J Antimicrob Chemother. 2016;71(8):2139-42. doi: 10.1093/jac/dkw145.
12. Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L, Galanakis C, et al. Food reservoir for Escherichia coli causing urinary tract infections. Emerg Infect Dis. 2010;16(1):88-95. doi: 10.3201/eid1601.091118.
13. Fibke CD, Croxen MA, Geum HM, Glass M, Wong E, Avery BP, et al. Genomic Epidemiology of Major Extraintestinal Pathogenic Escherichia coli Lineages Causing Urinary Tract Infections in Young Women Across Canada. Open Forum Infect Dis. 2019;6(11):ofz431. doi: 10.1093/ofid/ofz431.
14. Elankumaran P, Browning GF, Marenda MS, Reid CJ, Djordjevic SP. Close genetic linkage between human and companion animal extraintestinal pathogenic Escherichia coli ST127. Current Research in Microbial Sciences. 2022;3:100106. doi: 10.1016/j.crmicr.2022.100106.
15. Elankumaran P, Cummins ML, Browning GF, Marenda MS, Reid CJ, Djordjevic SP, et al. Genomic and Temporal Trends in Canine ExPEC Reflect Those of Human ExPEC. Microbiology Spectrum. 2022;10(3):e01291-22. doi: 10.1128/spectrum.01291-22.
16. Reid CJ, Cummins ML, Börjesson S, Brouwer MSM, Hasman H, Hammerum AM, et al. A role for ColV plasmids in the evolution of pathogenic Escherichia coli ST58. Nature Communications. 2022;13(1):683. doi: 10.1038/s41467-022-28342-4.
17. Forde BM, Roberts LW, Phan M-D, Peters KM, Fleming BA, Russell CW, et al. Population dynamics of an Escherichia coli ST131 lineage during recurrent urinary tract infection. Nature Communications. 2019;10(1). doi: 10.1038/s41467-019-11571-5.
18. Djordjevic SP, Stokes HW, Roy Chowdhury P. Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. Front Microbiol. 2013;4:86. doi: 10.3389/fmicb.2013.00086.
19. Jarocki VM, Heß S, Anantanawat K, Berendonk TU, Djordjevic SP. Multidrug-Resistant Lineage of Enterotoxigenic Escherichia coli ST182 With Serotype O169:H41 in Airline Waste. Front Microbiol. 2021;12:731050. doi: 10.3389/fmicb.2021.731050.
20. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. Microbiol Mol Biol Rev. 1999;63(3):507-22. doi: 10.1128/mmbr.63.3.507-522.1999.
21. 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. Microb Genom. 2017;3(12). doi: 10.1099/mgen.0.000143.
22. 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. Microb Genom. 2019;5(2). doi: 10.1099/mgen.0.000250.
23. Hoang T, da Silva AG, Jennison AV, Williamson DA, Howden BP, Seemann T. AusTrakka: Fast-tracking nationalized genomics surveillance in response to the COVID-19 pandemic. Nat Commun. 2022;13(1):865. doi: 10.1038/s41467-022-28529-9.

Appendixes

Appendix 1

Spreadsheet containing metadata for Chapter 4.

Accessible via link: https://figshare.com/s/8bc3d4afd6a0734cc9c9

Appendix 2

Pairwise SNP matrix of *E. coli* ST73 with all SNP counts of more than 100 removed.

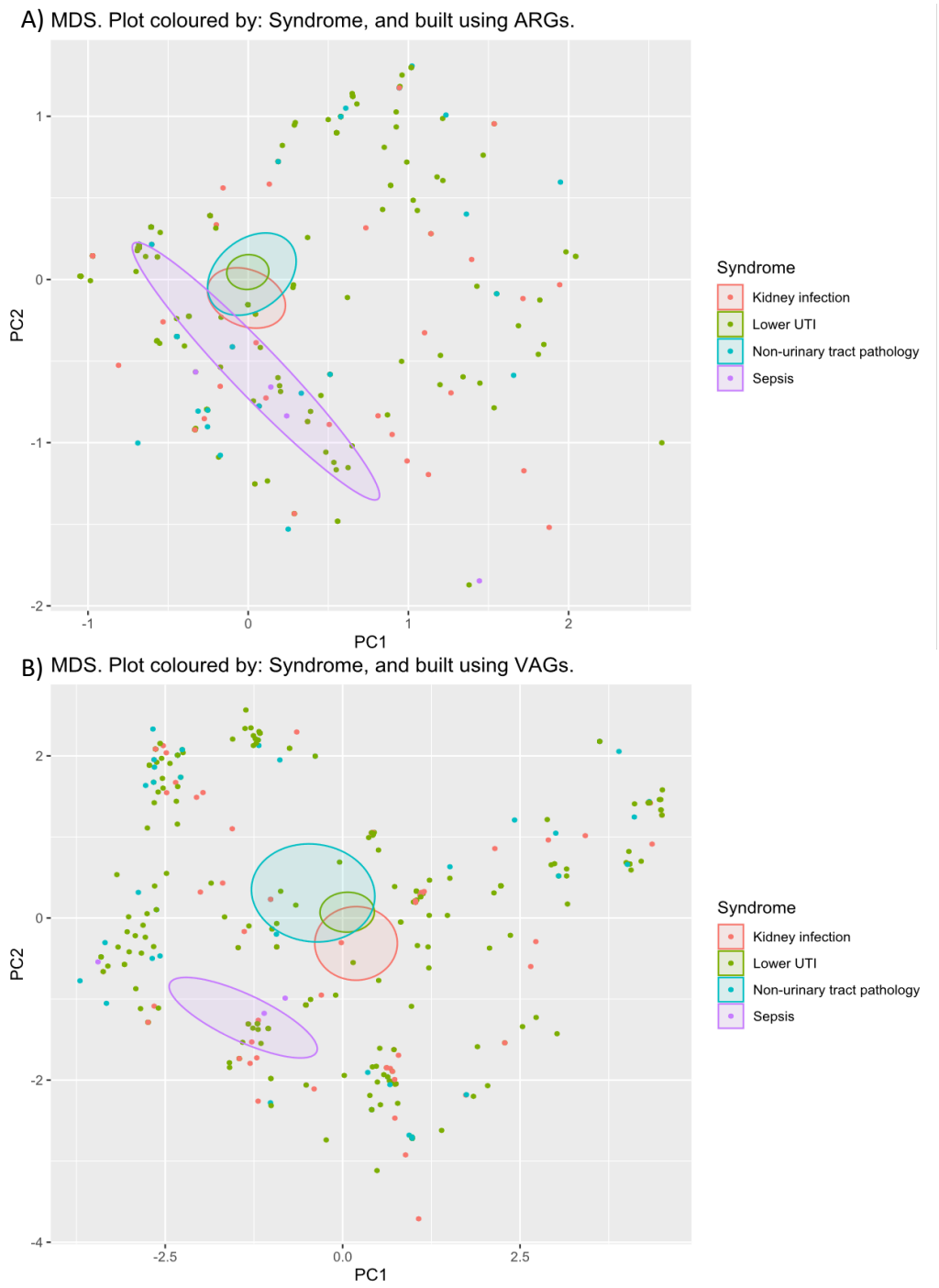
Accessible via link: https://figshare.com/s/68e8e2aa7e67448bed68

Appendix 3

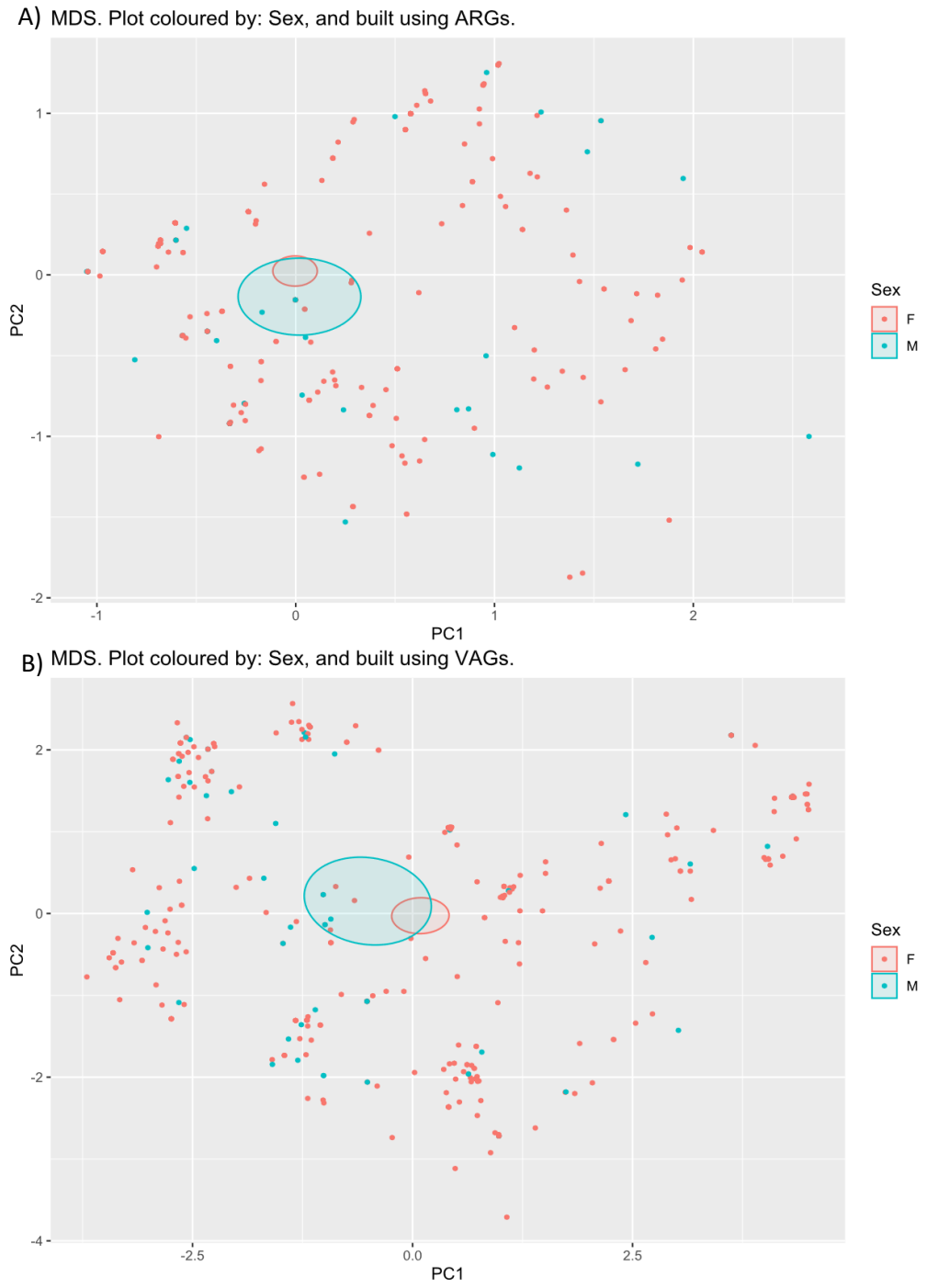
Gene screening results for Chapter 4.

Accessible via link: https://figshare.com/s/fc55dd4bf21ed28e9440

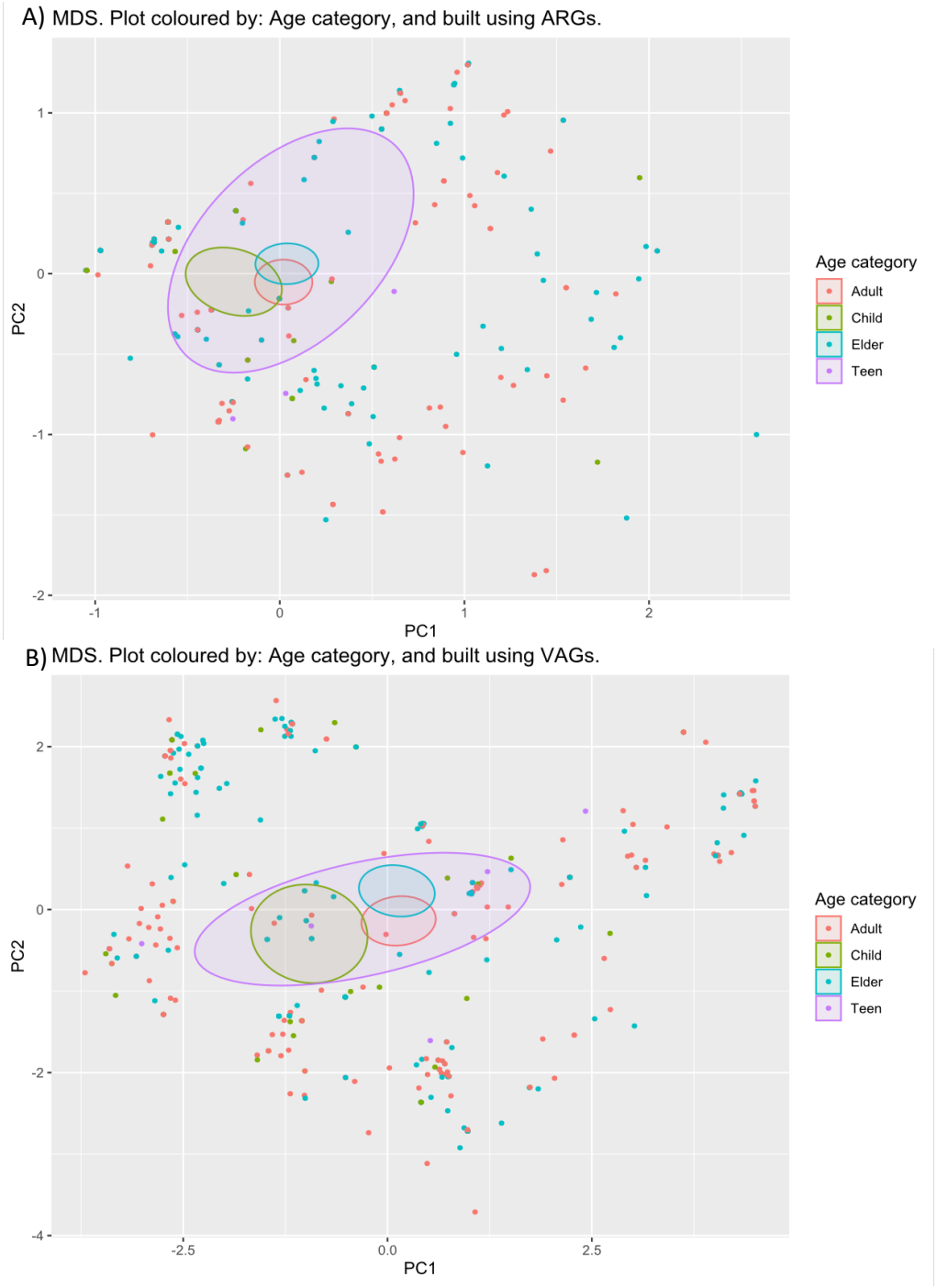
Appendix 4



Appendix 4.1: MDS analysis for A) ARG content and B) VAG content coloured by uro-clinical syndrome. All syndromes which are not related to UTIs are placed in the "Non-urinary tract pathology" group. Ellipsis represents confidence around the group mean points.

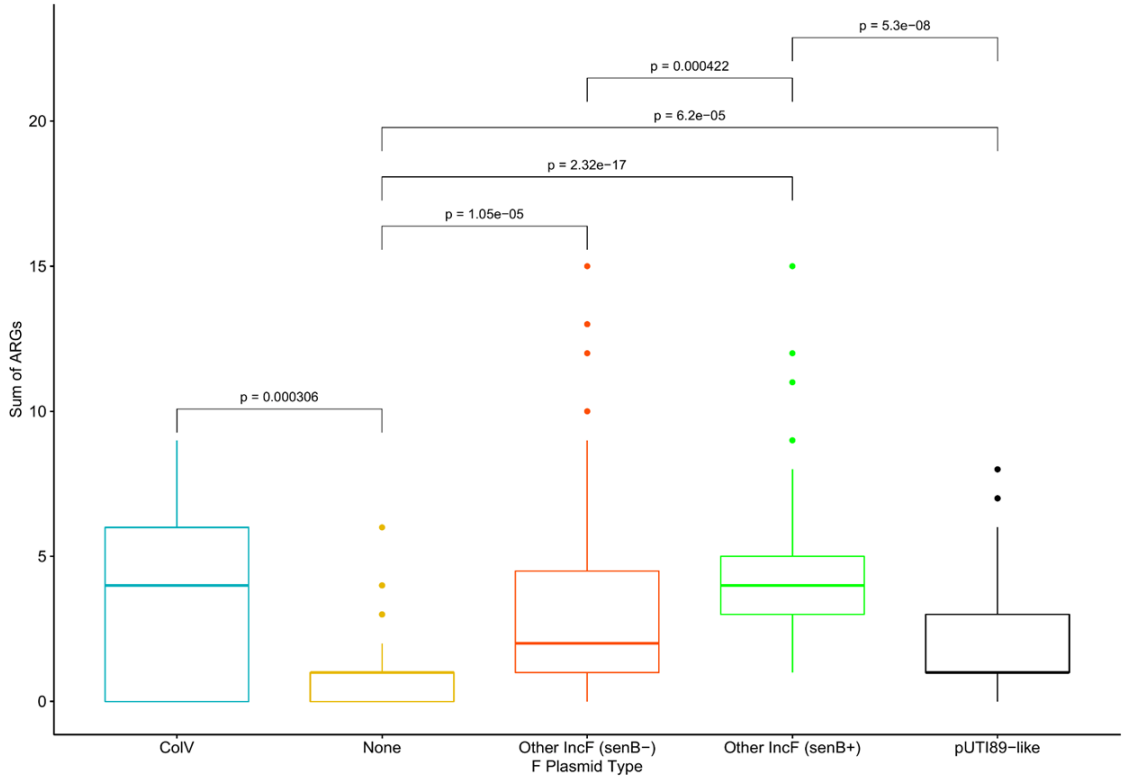


Appendix 4.2: MDS analysis for A) ARG content and B) VAG content coloured by sex (F: female, M: male). Ellipsis represents confidence around the group mean points.

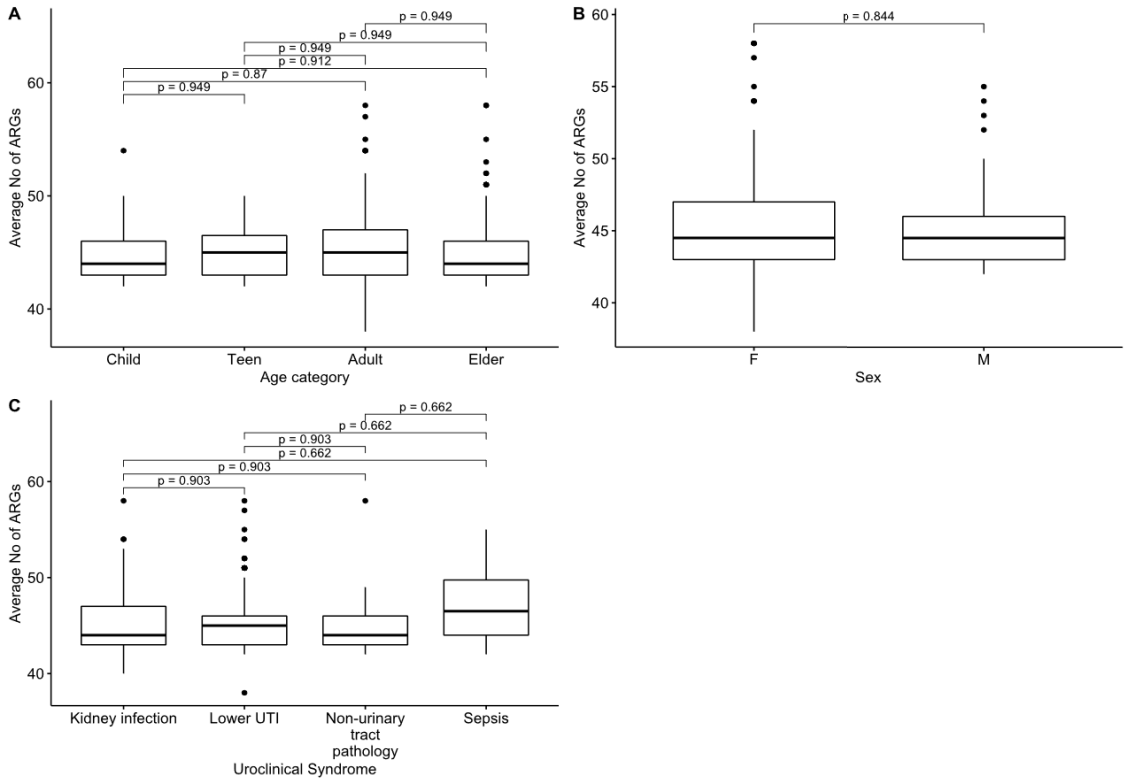


Appendix 4.3: MDS analysis for A) ARG content and B) VAG content coloured by age categories (Child: 0-11 yrs, Teen: 12-17 yrs, Adult: 18-64 yrs, Elder: 64+ yrs). Ellipsis represents confidence around the group mean points.

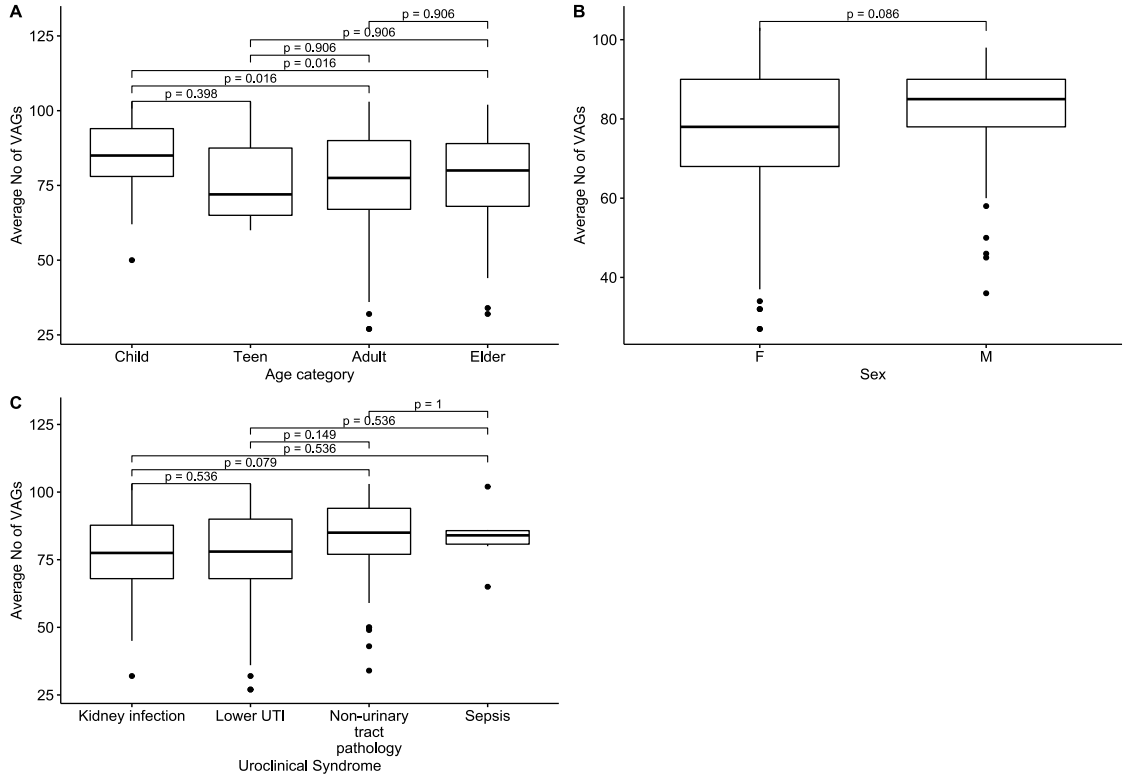
Appendix 5



Appendix 5.1: Boxplots comparing total ARG associated with HGT counts by different IncF plasmid types. Only $p < 0.05$ are shown, Pairwise Wilcoxon test with Benjamini–Hochberg adjusted p-values.



Appendix 5.2: Boxplots comparing total number of ARGs associated with A) Age categories (Child: 0-11 yrs Teen: 12-17 yrs, Adult: 18-64 yrs, Elder: 64+ yrs), B) Sexes (F: female, M: male) and C) Uroclinical syndromes, all syndromes which are not related to UTIs are placed in the "Non-urinary tract pathology" group. Pairwise Wilcoxon test with Benjamini–Hochberg adjusted p-values.



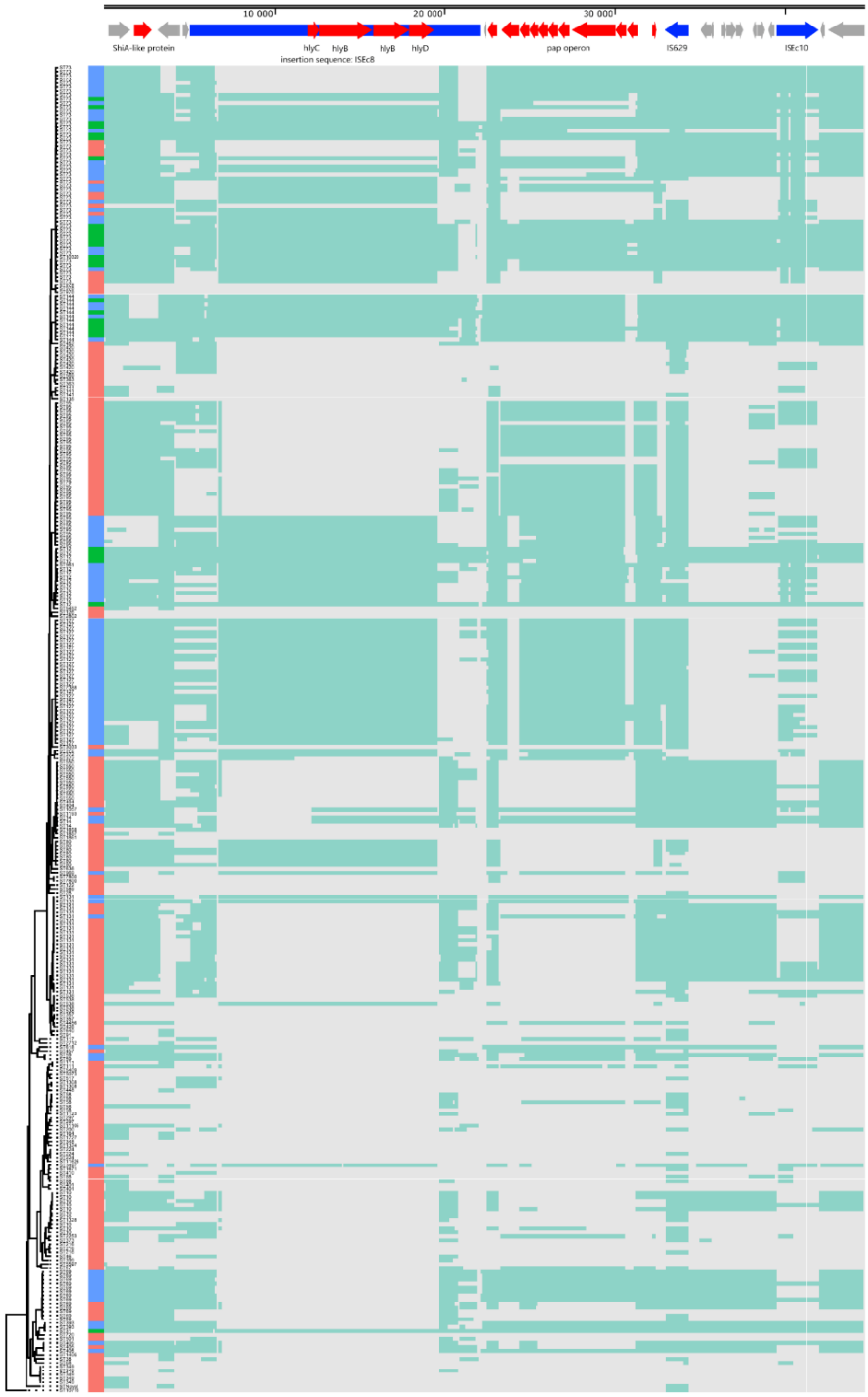
Appendix 5.3: Boxplots comparing total number of VAGs associated with A) Age categories (Child: 0-11 yrs, Teen: 12-17 yrs, Adult: 18-64 yrs, Elder: 64+ yrs), B) Sexes (F: female, M: male) and C) Uroclinical syndromes, all syndromes which are not related to UTIs are placed in the "Non-urinary tract pathology" group. Pairwise Wilcoxon test with Benjamini–Hochberg adjusted p-values.

Appendix 6

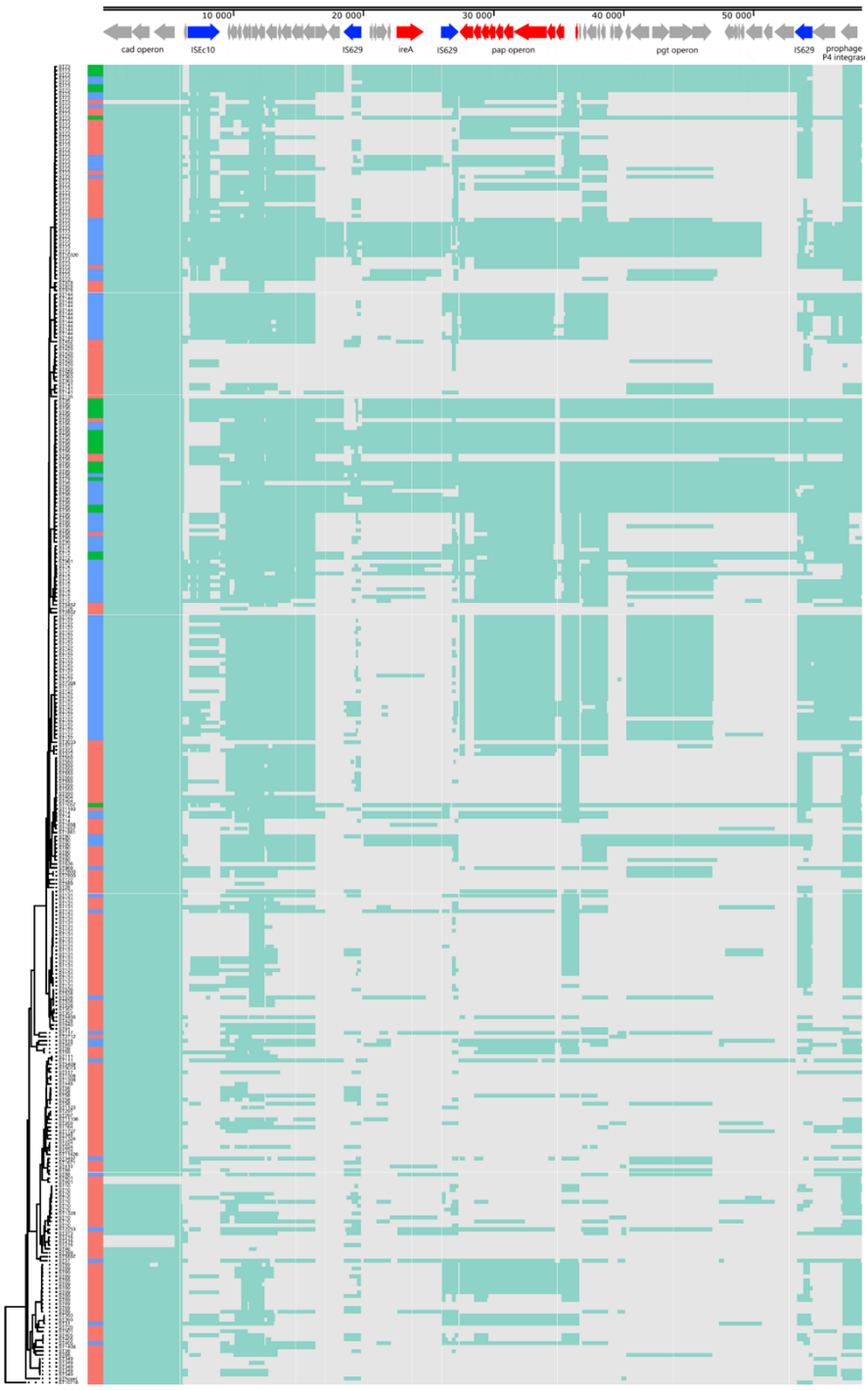
Result of GWA analysis performed by Scoary.

Accessible via link: https://figshare.com/s/9ccefaa13437942b5a02

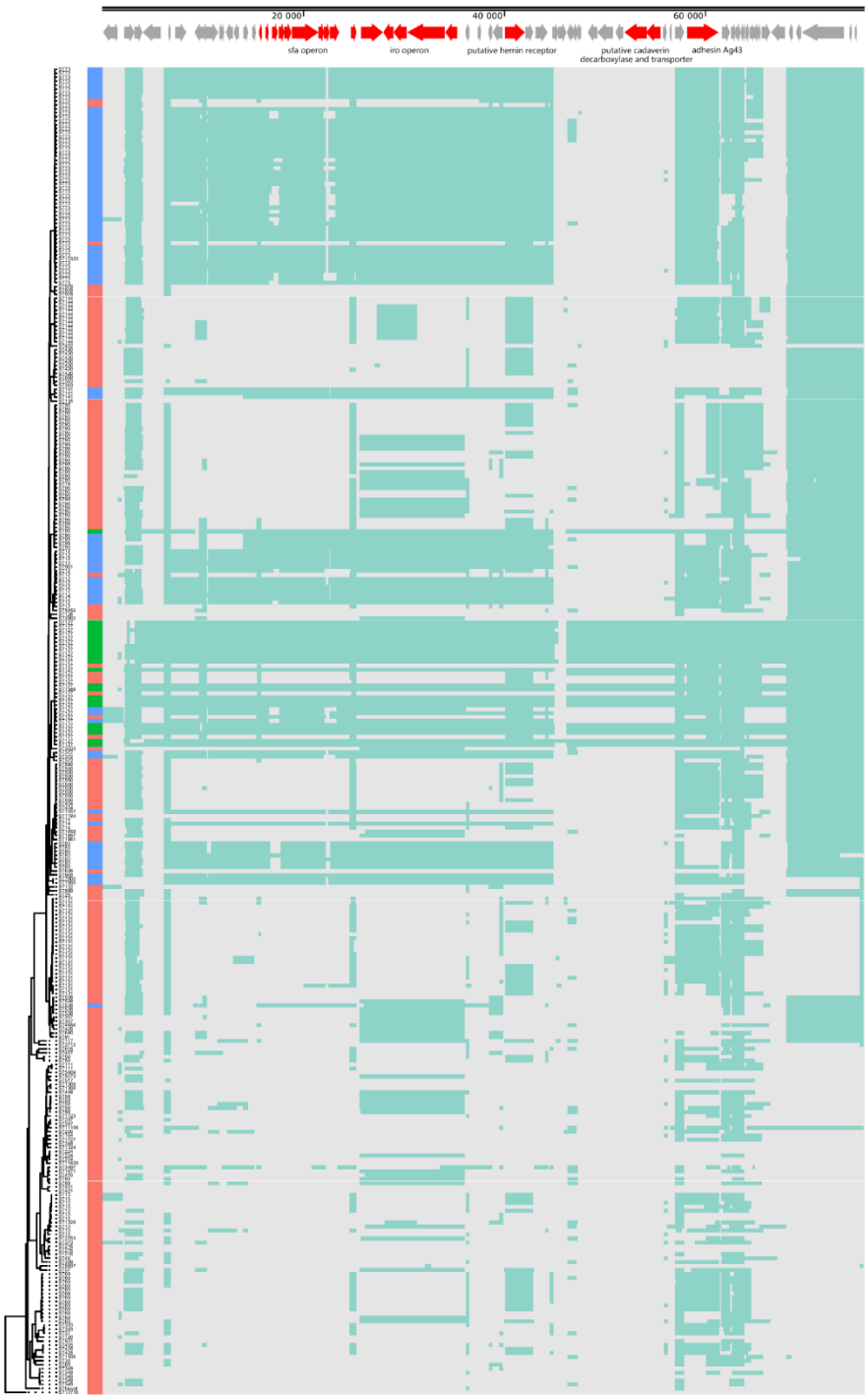
Appendix 7



Appendix 7A: Mapping of short-reads indicating the presence of PAI-I. Clustering of rows is based on the core genome phylogenetic tree (present on the left side). Isolate names replaced with corresponding ST. Presence/partial/absence of the PAI is presented as Green/Blue/Red box near the isolates.



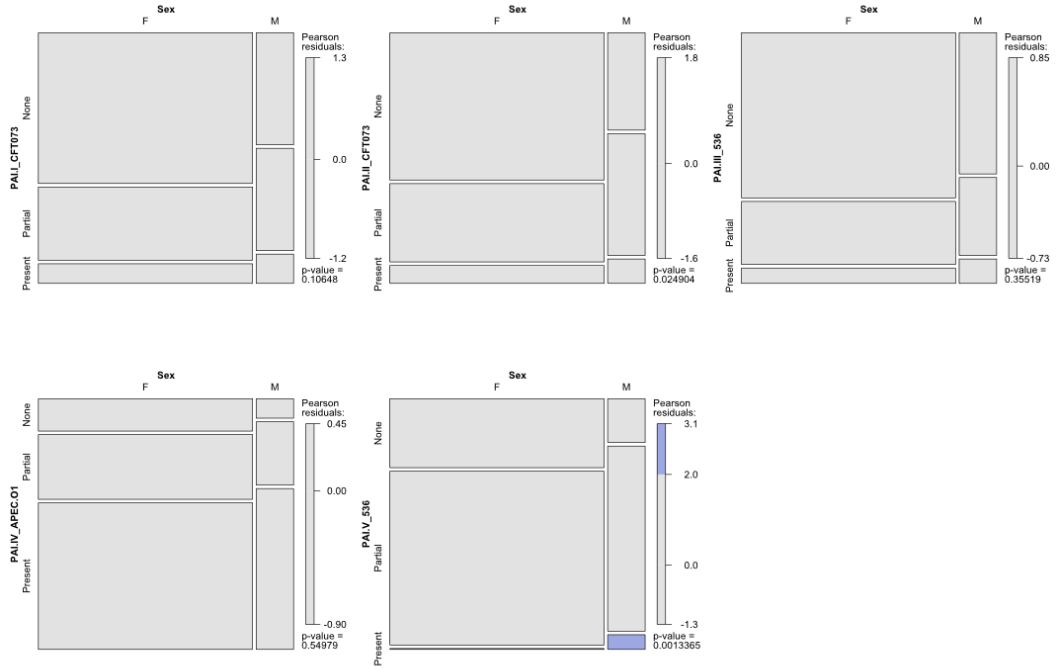
Appendix 7B: Mapping of short-reads indicating the presence of PAI-II. Clustering of rows is based on the core genome phylogenetic tree (present on the left side). Isolate names replaced with corresponding ST. Presence/partial/absence of the PAI is presented as Green/Blue/Red box near the isolates.



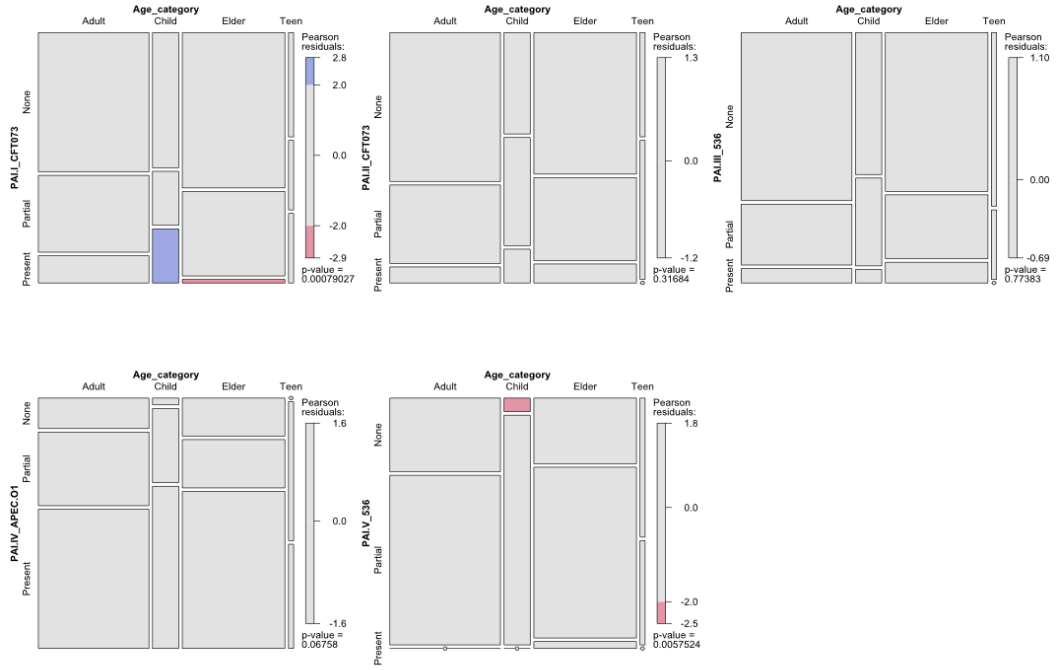
Appendix 7C: Mapping of short-reads indicating the presence of PAI-III. Clustering of rows is based on the core genome phylogenetic tree (present on the left side). Isolate names replaced with corresponding ST. Presence/partial/absence of the PAI is presented as Green/Blue/Red box near the isolates.

Appendix 7D: Mapping of short-reads indicating the presence of PAI-IV. Clustering of rows is based on the core genome phylogenetic tree (present on the left side). Isolate names replaced with corresponding ST. Presence/partial/absence of the PAI is presented as Green/Blue/Red box near the isolates.

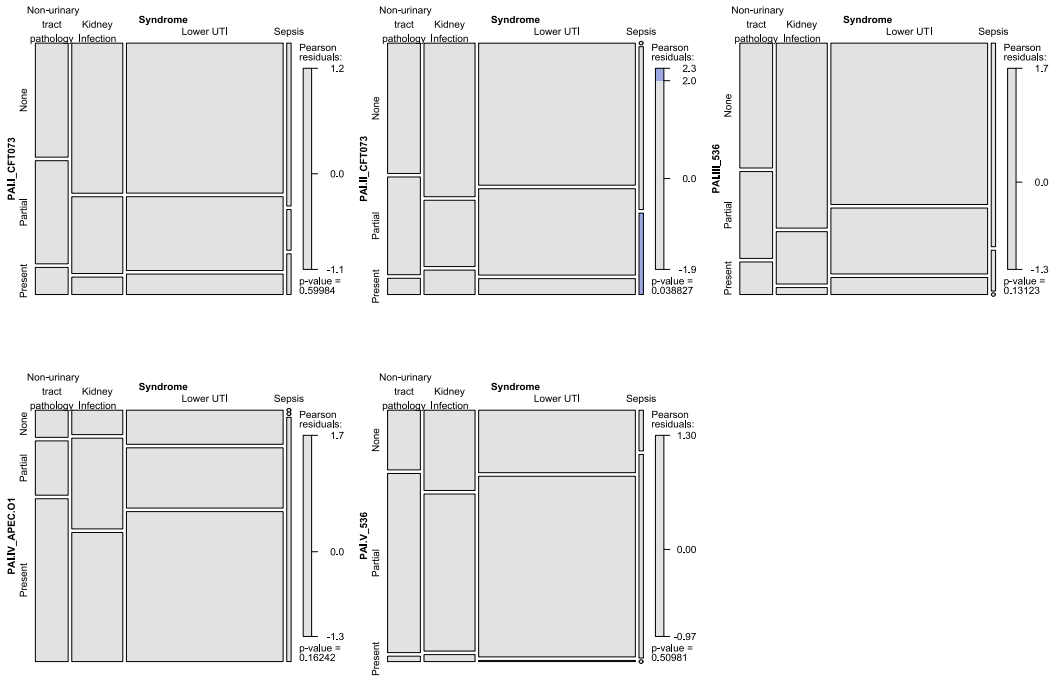
Appendix 7E: Mapping of short-reads indicating the presence of PAI-V. Clustering of rows is based on the core genome phylogenetic tree (present on the left side). Isolate names replaced with corresponding ST. Presence/partial/absence of the PAI is presented as Green/Blue/Red box near the isolates.



Appendix 7F: A mosaic plots showing cross-sectional distribution of PAIs and sexes. Coloured according to Pearson residuals, Blue – higher than expected, Red – lower than expected. p-value from X-square test.



Appendix 7G: A mosaic plots showing cross-sectional distribution of PAIs and age categories. Coloured according to Pearson residuals, Blue – higher than expected, Red – lower than expected. p-value from X-square test.



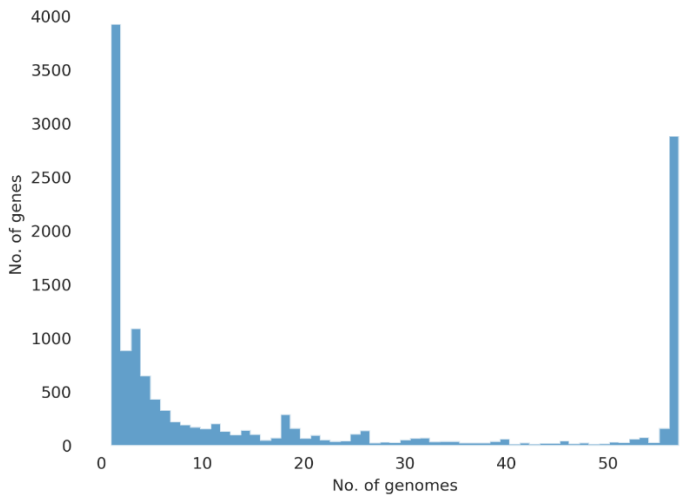
Appendix 7H: A mosaic plots showing cross-sectional distribution of PAIs and uroclinical syndromes. Aall syndromes which are not related to UTIs are placed in the "Non-urinary tract pathology" group. Coloured according to Pearson residuals, Blue – higher than expected, Red – lower than expected. p-value from X-square test.

Appendix 8

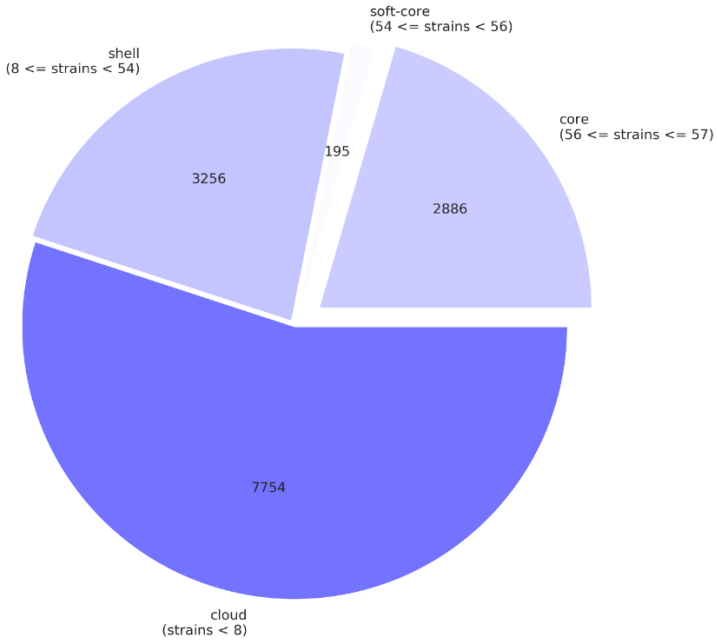
Statistics of genome assemblies for Chapter 5.

Accessible via link: https://figshare.com/s/ea6aa2972f995728792b

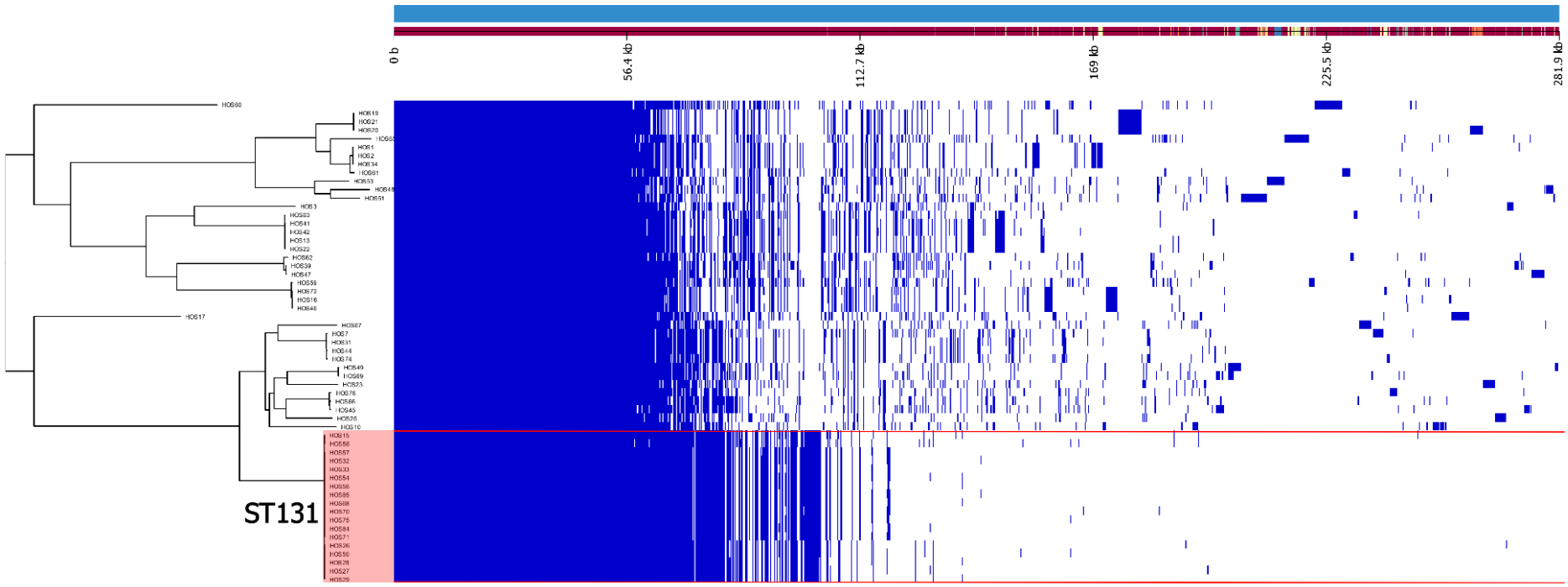
Appendix 9



Appendix 9A: Frequency of genes versus the number of genomes depicting U shape graph with high number of singlets and core genes.



Appendix 9B: The pie chart of the breakdown of genes and the number of isolates they are presented in. Representing relative abundance of each group of genes with vast majority to be cloud genes (genes which are present in less than 8 samples).



Appendix 9C: Pangenome gene presence/absence map highlighting similarities between EXPEC STs. All genomes annotated by Prokka, analysed by Roary, and visualised by Phandango

Appendix 10

Gene screening results and metadata table for Chapter 5.

Accessible via link: https://figshare.com/s/914dd72ae6f05fc0ef46

Appendix 11

ARGs to phenotype correlation heatmap.

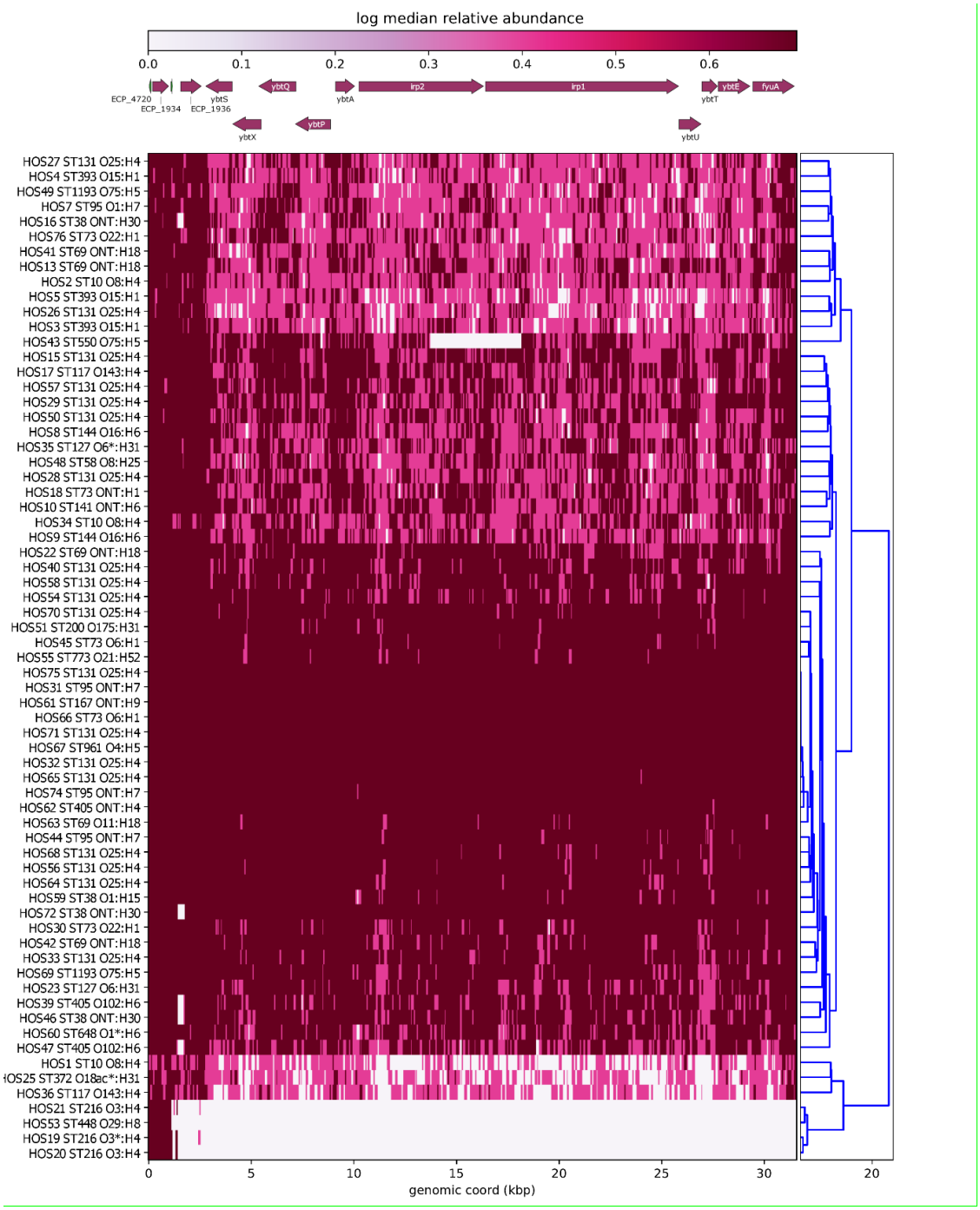
Accessible via link: https://figshare.com/s/74acb30f5773fa3a3ced

Appendix 12

VAGs and ARGs to phenotype correlation heatmap.

Accessible via link: https://figshare.com/s/4d5e1e09a375929a4ecc

Appendix 13



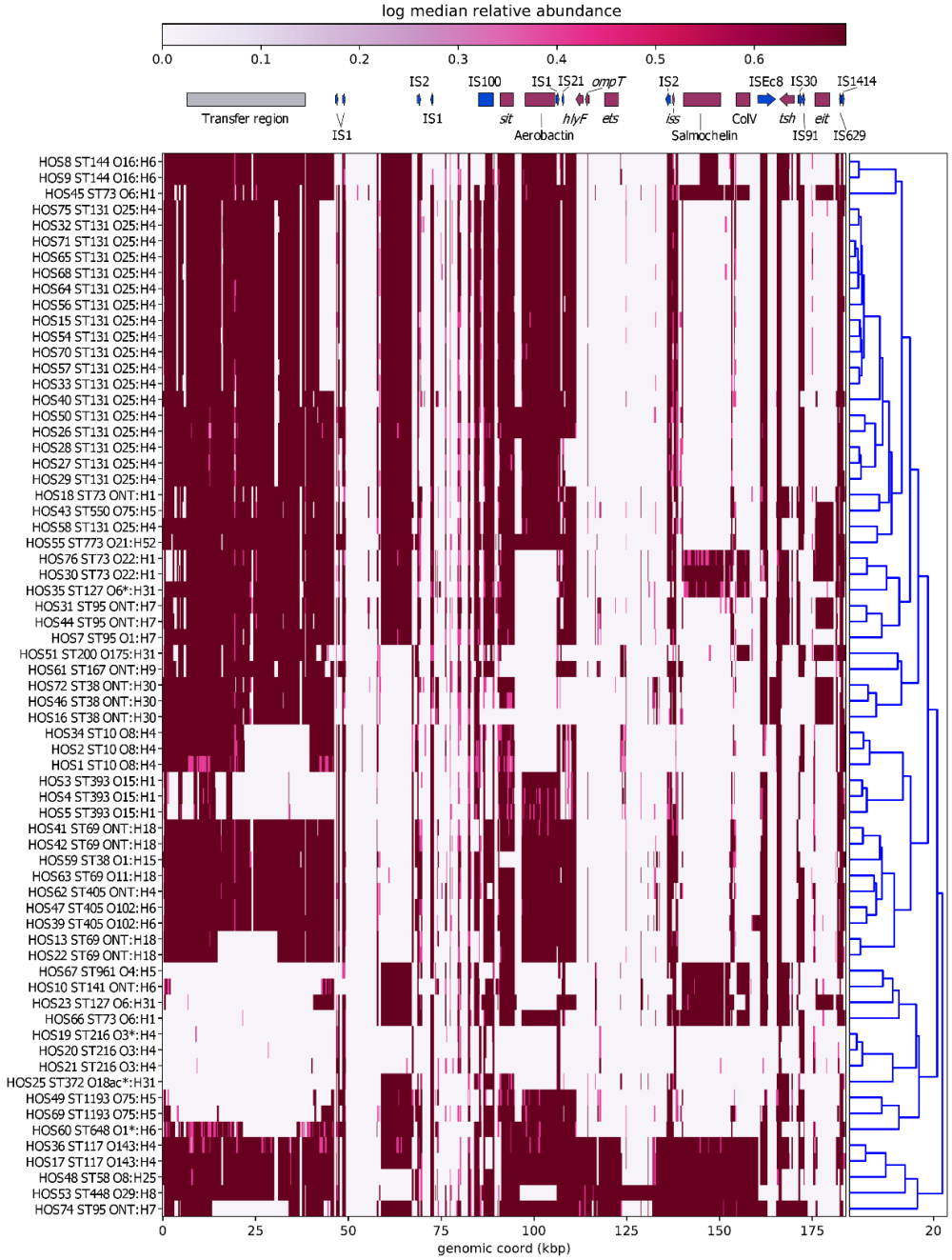
Appendix 13: Short-read mapping to PAI IV536/HPI. Tree to the left of heatmap were constructed by hierarchical clustering using Euclidean agglomeration method

Appendix 14

Blast search results for PAI IV536.

Accessible via link: https://figshare.com/s/f484db601505a59f0576

Appendix 15



Appendix 15: Short-read mapping to colV plasmid pAPEC-O2-ColV. Tree to the left of heatmap were constructed by hierarchical clustering using Euclidean agglomeration method

Appendix 16

SNP analysis of serial isolates. Coloured amino acid changes signify negative values for substitutions using BLOSUM64 matrix [reference]. Green = -1 or -2 Red = -3 or -4

Isolate	ST/Serotype	Diagnosis	Time difference from initial (days)	
HOS1	ST10/O8:H4	Cystitis	-	
HOS2	ST10/O8:H4	Pyelonephritis	63	
HOS34	ST10/O8:H4	Pyelonephritis	78	
SNP Analysis				
Gene	**SNP or variation**	**Amino acid change**	**Recurrent infection**	**Different Patient**
speC	G->T	H333Q	✓	✓
gspI	G->C	H67Q	✓	✓
irp1	T->G	D2717A	✓	✓
ybtS	C->G	P184A	✓	✓
	CA->GC	A189G	✓	✓
astA	G->T	G103V	✓	✓
aroC	C->T	G152E	✓	✓
menD	C->T	D412N	✓	✓
hyfD	T->G	L374V	✓	✓
fadH	C->A	Silent I98	✓	✓
astD	C->T	Silent A37	✓	✓
rnhB	C->A	R98L	✓	✓
hrpB	C->T	G644E	✓	✓
traC	G->T	G493W	✓	✓
	G->C	E494Q	✓	✓
recC	G->T	S794C	✓	✓
	G->T	H793N	✓	✓
mfd	G->T	A432D	✓	
zntA	A->G	V411A	✓	✓
ugpC	C->A	L297M	✓	✓
modB	T->G	Silent L208	✓	✓
lsrA	A->G	E296G	✓	✓
	G->C	Silent A306		✓
drrB	G->T	T99N	✓	✓
chemotaxis YjcZ-like protein	G->T	Silent P172	✓	✓
YdbH family protein	A->G	L103P	✓	
	G->A	Silent S96	✓	✓
mobA	T->G	S252A		✓
14 bp insertion between partial *btuC* and hypothetical	A-> ACCTCAGTTGATTTG		✓	
Between two tRNA-Gly	TAAG->CAAA		✓	
Between *yjjG* and *prfC*	A->G		✓	
Between *sodC2* and scaffold break	A->G			✓
Between *paaD* and contig break	C->T			✓

Between tRNA-Arg and scaffold break	G->A			✓
Isolate	**ST/Serotype**	**Diagnosis**	**Time difference from initial (days)**	**Total SNPs**
HOS3	ST393/O15:H1	Cystitis	-	-
HOS4	ST393/O15:H1	Pyelonephritis	284	13
HOS5	ST393/O15:H1	Pyelonephritis	412 (128 after recurrent 1)	19
SNP Analysis				
Gene	**SNP or variation**	**Amino acid change**	**Recurrent infection 1**	**Recurrent infection 2**
fecA	G->A	E124D	✓	
allC	A->G	I148T	✓	
galE	G->GA	Addition of E39 causing fusion with *capD*	✓	✓
fimH	G->A	A23V		✓
*rfaL**	T->TGGTTAC	Insertion of G and Y at 287		✓
ybiB	T->C	S227G	✓	
folC	A->T	Silent V93	✓	✓
argH	T->C	Silent A61	✓	✓
cysH	GCGCACAGCAGAA->G	Deletion of R, T, A and Q at 149-152	✓	
gpsA	C->T	T135M	✓	
udp	C->T	Q166STOP causing 9.9 kDa truncation		✓
mukF	A->C	F401V	✓	✓
	CGT->GTG	T394H	✓	✓
*ompF**	C->CT	Causing addition of 13.7 kDa		✓
BapA prefix-like domain-containing protein	C->G	STOP6042Y causing fusion with hypothetical protein	✓	✓
Autotransporter *	GAT->G	Deletion causing truncation of 39.3 kDa		✓
Between *yqjH* and tRNA-Met	G->C		✓	
Between *Alaa* and *LrhA*	T->C		✓	
Small contig, possibly within fibronectin type III protein	G->A			✓
	T->G			✓
	C->T			✓
Between *fimA* and *fimE*	T->TCC			✓
Between *gldA* and *yiiJF**	TTACC->ATACG			✓
	T->C			✓
	TC->GA			✓
	A->C			✓
Isolate	**ST/Serotype**	**Diagnosis**	**Time difference from initial (days)**	**Total SNPs**

HOS8	ST144/O16:H6	Cystitis	-	-
HOS9	ST144/O16:H6	Cystitis	0	5

SNP Analysis

Gene	SNP or variation	Amino acid change
tnpB	C->T	G11R
	C->T	Silent L116
cysJ	C->G	V371L
hofM	C->G	Silent A232
ves	C->G	G7R

Isolate	ST/Serotype	Diagnosis	Time difference from initial (days)	Total SNPs
HOS26	ST131/O25:H4	Cystitis	-	-
HOS27	ST131/O25:H4	Urosepsis	181	53
HOS28	ST131/O25:H4	Urosepsis	282 (101 after recurrent infection 1)	61
HOS29	ST131/O25:H4	Urosepsis	285 (3 after recurrent infection 2)	62
HOS50	ST131/O25:H4	Pyelonephritis	27 before initial presentation	21

SNP Analysis

Gene	SNP or variation	Amino acid change	Recurrent infection 1	Recurrent infection 2	Recurrent infection 3	Different Patient
kpsM	G->T	R18L	✓	✓	✓	✓
speC	C->T	A385	✓	✓	✓	✓
cyaA	T->C	V315A	✓	✓	✓	✓
narI	T->G	M128L	✓	✓	✓	✓
	G->C	A124G	✓	✓	✓	✓
iutA	C->T	G419D	✓	✓	✓	
*iucC**	A->G	Silent H38	✓	✓	✓	
	GGC->TGT	A25T	✓	✓	✓	
	C->A	Silent T16	✓	✓	✓	
	A->C	H12Q	✓	✓	✓	
	T->G	M9L	✓	✓	✓	
	C->A	E8D	✓	✓	✓	
	A->G	Silent C5	✓	✓	✓	
*sitC**	T->C	Silent K279	✓	✓	✓	
	C->T	Silent V230	✓	✓	✓	
	T->G	I177L	✓	✓	✓	
	G->A	Silent S115	✓	✓	✓	
	G->A	Silent G95	✓	✓	✓	
	A->G	Silent T86	✓	✓	✓	
	A->C	L85R	✓	✓	✓	
	T->A	Silent A75	✓	✓	✓	
	G->A	Silent L46	✓	✓	✓	

*sitB**	T->A	Silent A264	✓	✓	✓	
	C->T	Silent L234	✓	✓	✓	
	C->T	Silent T199	✓	✓	✓	
	G->A	Silent T171		✓	✓	
	C->G	Silent Q179	✓	✓	✓	
	C->T	D134N	✓	✓	✓	
	C->T	Silent Q121	✓	✓	✓	
	C->A	Q88D	✓	✓	✓	
	G->A	Silent L75	✓	✓	✓	
	T->C	Silent T71	✓	✓	✓	
	C->T	Silent G62	✓	✓	✓	
	GCC->ACT	G32S	✓	✓	✓	
*sitA**	T->C	Silent G110	✓	✓	✓	
	G->A	Silent A109	✓	✓	✓	
	C->A	Silent L96	✓	✓	✓	
	A->C	Silent P90	✓	✓	✓	
	A->G	Silent G74	✓	✓	✓	
	C->T	Silent A71	✓	✓	✓	
*flgK**	GCAATGGCGTCTACGTTTCTGGTGT GCA->G	Deletion of N, G, V, Y, V, S, G, V and Q at 55		✓	✓	
TerC family protein	G->A	Silent V36		✓	✓	
irp2	C->T	G201E		✓	✓	✓
moeA	G->T	R268L	✓	✓	✓	✓
galF	T->C	Silent R118				✓
aceK	C->T	Silent N550	✓			
prpE	T->C	Silent K587	✓			
nrdD	C->T	G389S	✓	✓	✓	✓
putative DNA-binding transcriptional regulator	T->G	E69D	✓	✓	✓	✓
dinG	T->A	I318L	✓	✓	✓	✓
uroporphyrinogen-III C-methyltransferase*	CGGTGCA->C	Deletion of P and A at 389			✓	
mdtK	T->C	Silent R302	✓	✓	✓	

hycC	T->G	S585A		✓	✓	✓
traC	A->G	S550P		✓	✓	
glp	C->A	Silent V107	✓	✓	✓	✓
ycdY	T->C	M58R	✓	✓	✓	✓
ag43/flu	A->G	Silent G752	✓	✓	✓	✓
sixA	A->G	Silent A117	✓	✓	✓	✓
yjbH domain-containing protein	G->A	Silent E52		✓	✓	
yaiW	T->C	Q24R		✓	✓	
gpH	G->T	Silent A716		✓	✓	✓
	G->C	R715G		✓	✓	✓
	A->C	C713G		✓	✓	✓
Hypothetical protein WLH_00908	G->T	G129V		✓	✓	
Between partial *mpi* and *dpK*	GGCGC->TGCAT		✓			
Between *AraH* and *otsB*	C->T		✓			
Between *wrbA* and MBL fold metallo-hydrolase	C->T		✓	✓	✓	✓
Between *SitC* and scaffold break	T->C		✓	✓	✓	
	A->T		✓	✓	✓	
Between replication protein rep and scaffold break	T->C			✓	✓	✓
Between *caiE* and scaffold break	A->C				✓	

Isolate	**ST/Serotype**	**Diagnosis**	**Time difference from initial (days)**	**Total SNPs**
HOS41	ST69/ONT:H18	Urosepsis	-	-
HO542	ST69/ONT:H18	Urosepsis		12

SNP Analysis

Gene	**SNP or variation**	**Amino acid change**	**Recurrent infection**
ybtT	C->A	R133L	
	C->G	E132Q	
	C->T	R128H	
lpxM	C->G	Silent L70	
	A->C	STOP73C causing fusion	

		with *ipxM*	
sipD	C->A	N4K	
cvrA	T->G	R337S	
hypothetical protein	A->G	Silent I170	
hypothetical protein	CA->AG	M1A	
Between *yjjV* and scaffold break	C->A		
Between tRNA-Lys and *nadA*	G->A		
Between DUF905 domain-containing protein and *parB*	A->G		
Between hypothetical protein and scaffold break	T->C		

Appendix 17

Assembly statistics for Chapter 6.

Accessible via link: https://figshare.com/s/9a5b265b5931435bb477

Appendix 18

Melbourne Veterinary Collection (MVC) isolates

Clinical specimens from canines were inoculated on Sheep Blood Agar and MacConkey Agar plates and incubated for 18-24 h hours at 37°C with 5% CO_2. Pure cultures of lactose fermenting organisms were identified as *E. coli* by conventional biochemistry (Indole, Methyl Red, Voges-Proskauer and Citrate). Isolates were stored at -80°C in a Protect beads cryogenic system (Technical Service Consultants, distributed by ThermoFisher scientific). All *E. coli* isolates from the MVC were freshly cultured onto LB agar plates and a single colony used to inoculate 5 mL of sterile LB media. After overnight culture 500 µL of 50% (v/v) glycerol was added and the stock was stored at -80°C. Total cellular DNA was extracted using the ISOLATE II Genomic DNA (Bioline) kit following the manufacturer's standard protocol for bacterial cells and stored by refrigerating at 4°C. Library preparation was done by the iThree Core Sequencing facility, University of Technology Sydney, following the adapted Nextera Flex library preparation kit process, Hackflex (1). Briefly, genomic DNA was quantitatively assessed using Quant-iT picogreen dsDNA assay kit (Invitrogen, USA). After tagmentation, DNA was amplified using the facility's custom designed i7 and i5 barcodes, with 12 cycles of PCR. After library amplification, 3 µL of each library was pooled into a library pool. The pool was cleaned using SPRIselect beads (Beckman Coulter, USA) following the Hackflex protocol. The final pool was sequenced on one lane of Illumina Novaseq S4 flow cell, 2x150 bp at Novogene (Singapore).

Orange Base Hospital (OBH) isolates

Human urine specimens were collected and sequenced as described previously (2). Briefly, urine specimens were collected, and semi-quantitative culture performed on horse blood, MacConkey and chromogenic agars, followed by conventional identification. DNA for WGS was extracted using the ISOLATE II genomic DNA kit (Bioline) and stored at −20°C. Short-read sequencing was performed using an Illumina HiSeq 2500 v4 sequencer in rapid PE150 mode. In this study, OBH isolates consist of both published and unpublished genomes. Unpublished genomes were HOS77-99 (n=23), and previously published isolates were HOS15-75 (n=19) (2).

Australian silver gull isolates

Isolates sourced from silver gulls were collected and sequenced as described previously (3). Briefly, isolates were sourced from cloacal swabs of silver gulls. Genomic DNA for WGS was isolated using NucleoSpin tissue kit (Macherey-Nagel GmbH & Co.). DNA libraries were prepared using Nextera XT DNA sample preparation kit and sequenced on a NovaSeq (Illumina, San Diego, USA) platform.

South Australia environmental (SAE) isolates

Raw influent samples (in triplicate) were collected from a South Australian municipal Wastewater Treatment Plant (hereafter referred to as WWTP). WWTP serves ≈700,000 inhabitants and receives around 175 ML/day of low-to-medium organic strength sewage with a large industrial/commercial component, as well as residential, and hospital sources (Table 1). Raw influent samples were collected in August, October and December 2019 in sterile polypropylene containers and stored on ice during transport to the laboratory (<2 h). On the day of collection raw influent samples were serially diluted and 500 µL from 2-3 consecutive 10-fold serial dilutions were plated in triplicates on Oxoid Brilliance™ CRE Agar plates (Thermo Fisher Scientific Australia, Adelaide, SA). Cultures were incubated at 37°C and 44°C. 44°C is *E. coli* recommended incubation temperature (4). The incubation at 37°C was chosen to reduce curing of plasmids harboring ARGs. Using sterilized inoculating loops, single colonies growing on CRE Agar were plated on Plate Counting Agar (PCA; Thermo Fisher Scientific Australia, Adelaide, SA). PCA cultures were incubated at 37°C and 44°C for 18-24 h, or until sufficient *E. coli* growth occurred. *E. coli* were identified with Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS, detailed protocol described in the supplementary information) and preserved in glycerol stocks (40% v/v) at -80°C. DNA from MALDI-TOF MS identified *E. coli* colonies was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Sydney, Australia) according to the manufacturer's instructions. Nucleic acid quality (i.e. 260/280 ratio) was measured with Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific Australia, Adelaide, SA). DNA extracts were quantified using a Quant-iT HS ds-DNA assay kit in a Qubit™ fluorometer (Invitrogen, Carlsbad, CA, USA), quality checked by 1.5% agarose gel electrophoresis and stored at −20°C until sequencing. WGS was performed as described previously (5). Briefly, WGS was performed on the Illumina NextSeq 500 platform using a modified Nextera low input tagmentation approach. Genomic DNA was normalized to 0.5 ng µL−1 with 10 mM Tris-HCl before the library preparation. The pooled library was run at a final concentration of 1.8 pM on a mid-output flow cell following Illumina recommended denaturation and loading parameters. Data was uploaded to Basespace (www.basespace.illumina.com), where the raw data was converted to FASTQ files for each sample.

Table 1. Characteristics of the wastewater treatment plant, showing average raw sewage properties, key process operating conditions and final effluent water quality.

WWTP	Pop. served[a]	PE organic load[b]	Sewage flow[c, *]	COD[d]	BOD_5[e]	TDS[f]	Bioreactor MLSS[g, *]	UV dose[h]	UVT[i]	Chlorine dose	Effluent free chlorine	Effluent total P	Effluent total N
	(*n*)	(*n*)	(ML/day)	(mg/L)	(mg/L)	(mg/L)	(g/L)	(mJ/cm²)	(%)	(mg/L)	(mg/L)	(mg/L)	(mg/L)
	695,630	1,150,000	175 (±15)	857	355	1,280	4	n/a	n/a	11[j]	1.1[j]	0.05	7.5

[a] Based-on census data provided by the managing water authority; [b] Population equivalent organic load calculated based on 120 grams COD/person/day (PE_{COD120}); [c] Average daily sewage flow rates given ± 1 standard deviation; [d] Chemical oxygen demand; [e] Five-day biochemical oxygen demand; [f] Total dissolved solids; [g] Mixed liquor suspended solids; [h] 254 nm ultraviolet (UV) light dose; [i] Ultraviolet light transmittance (UVT); [j] Chlorine data relate to media filtration plant; [*] Average parameters based on data from period July 2016–June 2017.

Sydney Adventist Hospital (SAH) isolates

Isolates with prefix SAH were recovered from patients presenting at the Sydney Adventist Hospital between 2009-2011. Isolates SAH2009_1, SAH2009_7, SAH2009_47, SAH2009_51, SAH2009_54, SAH2010_66 and SAH2011_90 were obtained from mid-stream urine, SAH2009_20 and SAH2009_37 were derived from catheter urine, and SAH2011_76 was isolated from mucosalivary sputum. Genomic DNA was isolated and sequenced on an Illumina HiSeq 2500 as previously described (6). Isolate SAH2009_36 was also from Sydney Adventist Hospital but has been previously published (6).

Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS)

For protein extraction, fresh bacterial isolates (24h old) were resuspended in 1 mL 70% ethanol, vortexed for 1 min, and centrifuged at 13,000 rpm for 2 min. The supernatant was removed completely, and the pellet re-dissolved with 5 µL of 70% formic acid (FA) (Baker; 90% stock) and 5 µL acetonitrile (ACN, LC-MS Grade, Merck). After 2 min centrifugation at 13,000 rpm, 1 µL of supernatant was spotted onto the target plate and left to dry. The sample was overlaid with α-cyano-4-hydroxycinnamic acid (HCCA) (1 µL) matrix (10 mg mL−1) and allowed to crystallise at room temperature. One µL Bacterial test standard (BTS, Bruker Daltonics) in 50% (v/v) ACN (LC-MS Grade, Merck) containing 2.5% (v/v) trifluoroacetic acid (TFA) (LC-MS Grade; Thermo Fisher Scientific) was spotted, left to dry and overlaid with HCCA for calibration. MALDI-TOF MS analysis was acquired on an autoflexTM speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) operated in linear positive mode under MALDI Biotyper 3.0 Real-time Classification (version 3.1, Bruker Daltonics) and FlexControl (version 3.4, Bruker Daltonics) software. Spectra were acquired in the mass range of 2000 to 20 000 Da with variable laser power, and a total of 1200 sum spectra were collected in 40 shot steps. The sample spectra were identified against an MSP database library (5989 MSP entries). Identification scores of 2.300–3.000 indicate highly

probable species identification, scores of 2.000–2.299 indicate secure genus identification and probable species identification, scores of 1.700–1.999 indicate probable genus identification, and a score of ≤1.699 indicates that the identification is not reliable.

References:

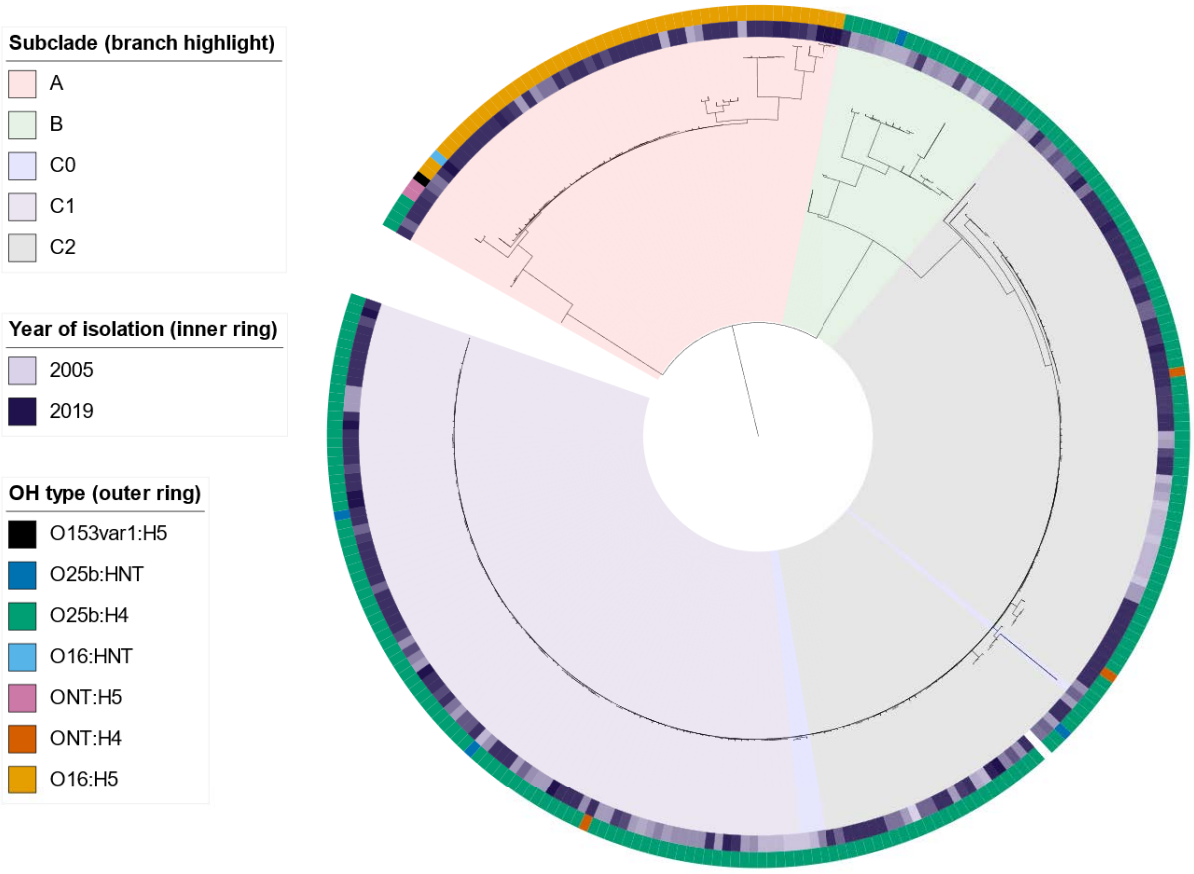
1. Gaio D, To J, Liu M, Monahan L, Anantanawat K, Darling AE. Hackflex: low cost Illumina sequencing library construction for high sample counts. [Preprint; cited 2021 Apr 20]. bioRxiv. 2019:779215. doi: 10.1101/779215.
2. Li D, Reid CJ, Kudinha T, Jarocki VM, Djordjevic SP. Genomic analysis of trimethoprim-resistant extraintestinal pathogenic Escherichia coli and recurrent urinary tract infections. Microbial Genomics. 2020. doi: 10.1099/mgen.0.000475.
3. Nesporova K, Wyrsch ER, Valcek A, Bitar I, Chaw K, Harris P, et al. Escherichia coli Sequence Type 457 Is an Emerging Extended-Spectrum-β-Lactam-Resistant Lineage with Reservoirs in Wildlife and Food-Producing Animals. Antimicrobial Agents and Chemotherapy. 2020;65(1). doi: 10.1128/aac.01118-20.
4. Marano RBM, Fernandes T, Manaia CM, Nunes O, Morrison D, Berendonk TU, et al. A global multinational survey of cefotaxime-resistant coliforms in urban wastewater treatment plants. Environ Int. 2020;144:106035. doi: 10.1016/j.envint.2020.106035.
5. Foster-Nyarko E, Alikhan NF, Ravi A, Thilliez G, Thomson NM, Baker D, et al. Genomic diversity of Escherichia coli isolates from non-human primates in the Gambia. Microb Genom. 2020;6(9). doi: 10.1099/mgen.0.000428.
6. Reid CJ, McKinnon J, Djordjevic SP. Clonal ST131-H22 Escherichia coli strains from a healthy pig and a human urinary tract infection carry highly similar resistance and virulence plasmids. Microb Genom. 2019;5(9). doi: 10.1099/mgen.0.000295.

Appendix 19

Metadata and gene screening results for Chapter 6.

Accessible via link: https://figshare.com/s/aaafd5a60218e0a0eeb8

Appendix 20



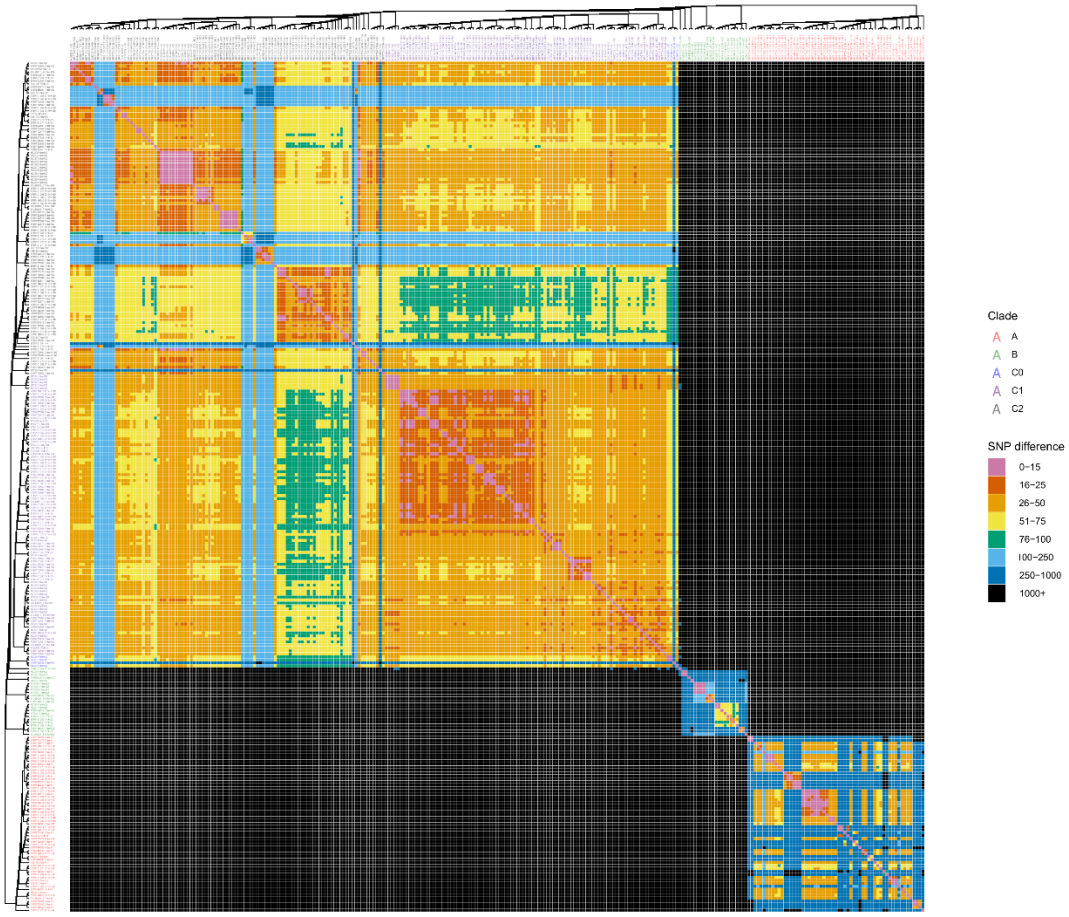
Appendix 20: Maximum-Likelihood SNP-based phylogenetic tree of Australian ST131. Inferred using FastTree2 and EC958 as a reference without recombination filtering applied. Sections of the tree coloured according to subclade (red = A, green = B, blue = C0, violet = C1 and black = C2). Tree scale bars represent the number of substitutions per site of alignment.

Appendix 21

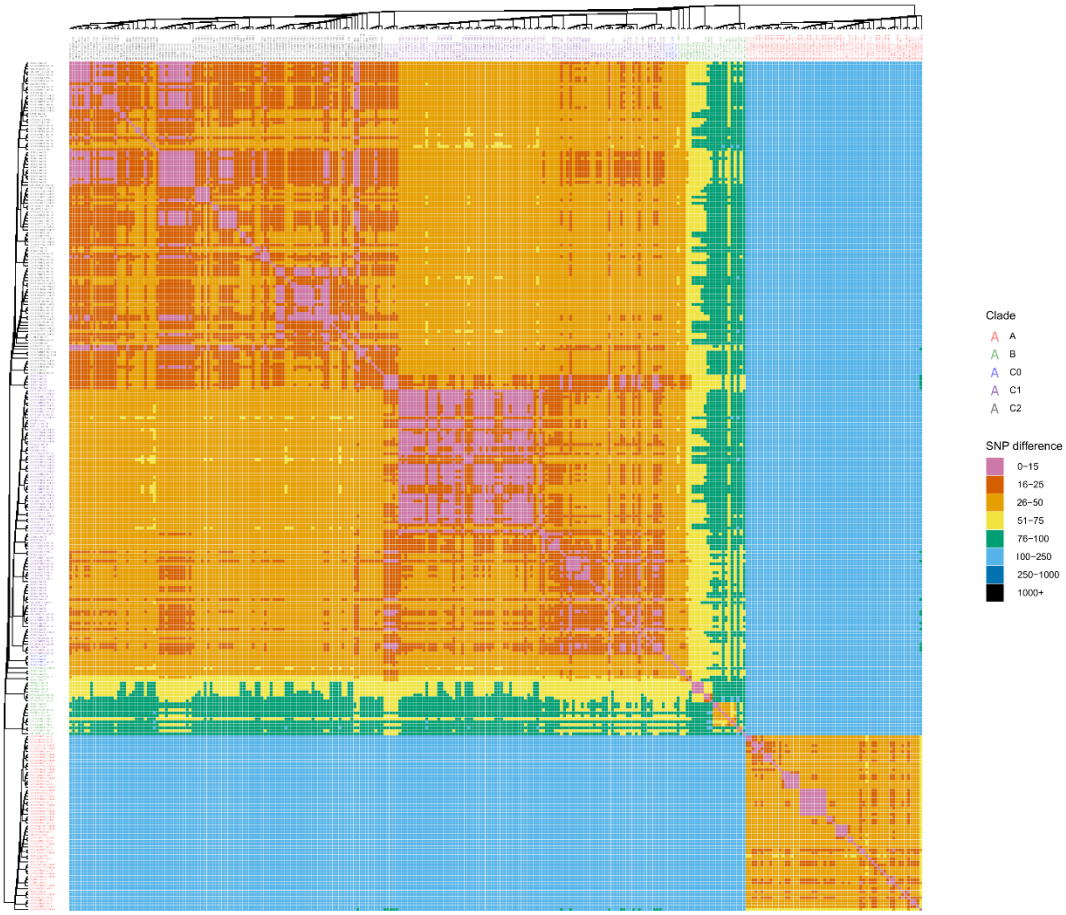
Gene presence/absence and GWA analysis for Chapter 6.

Accessible via link: https://figshare.com/s/ec94a33bdf43f8a2021e

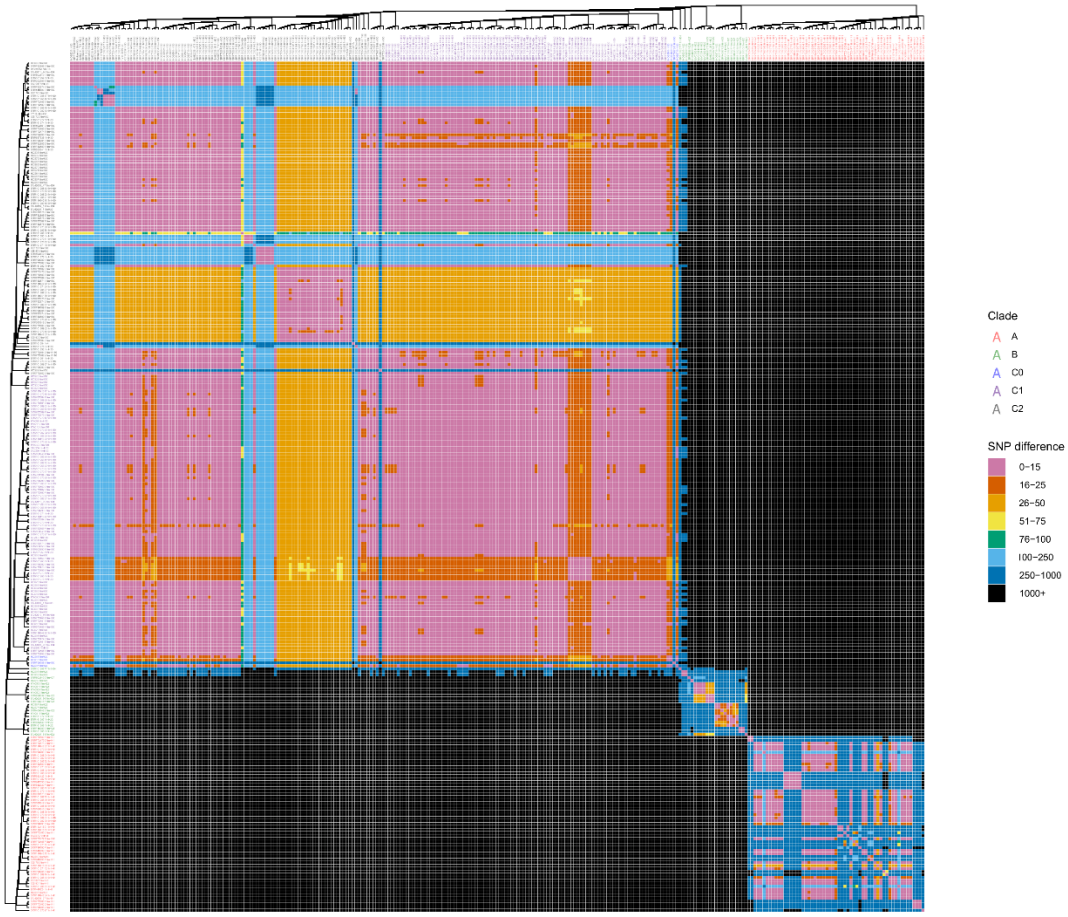
Appendix 22



Appendix 22A: Pairwise SNP distance heatmap, without recombination filtering.



Appendix 22B: Pairwise SNP distance heatmap, with recombination filtering.

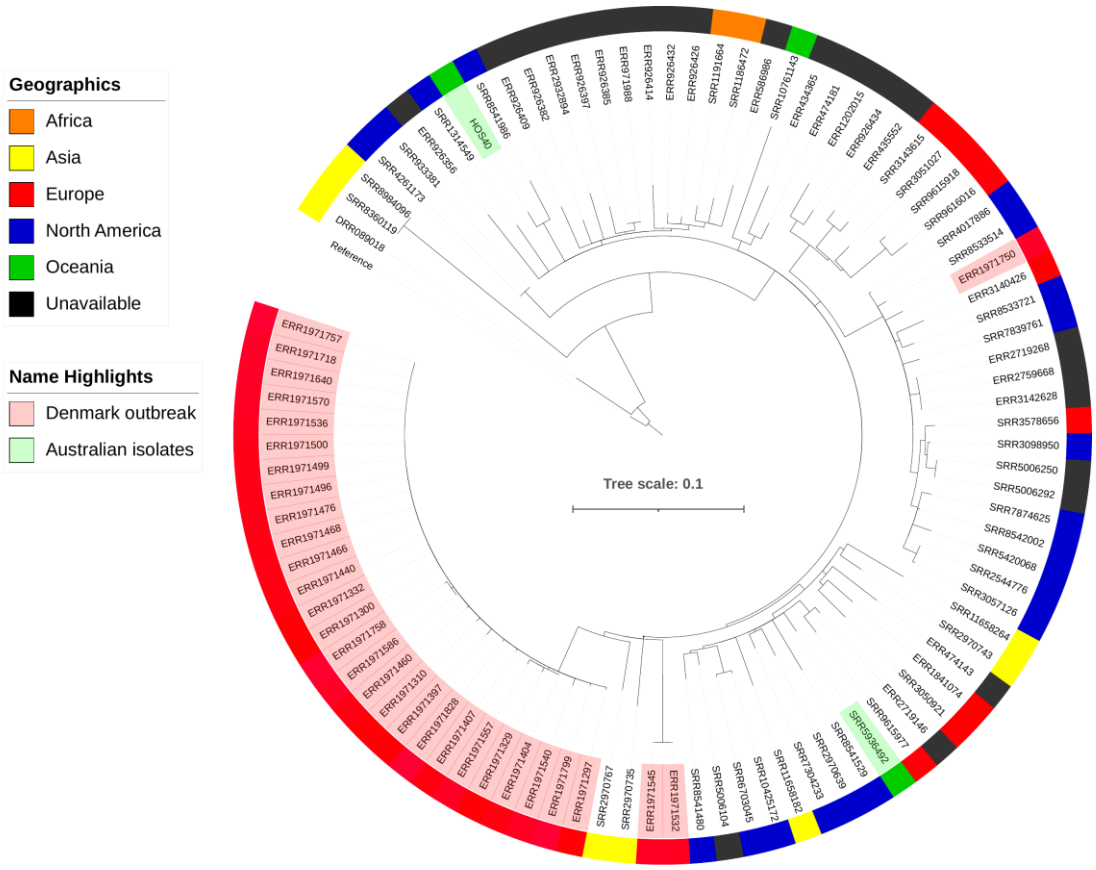


Appendix 22C: Pairwise SNP distance heatmap, providing SNPs due to recombination only.

Appendix 23

M27PP1 screening results.

Accessible via link: https://figshare.com/s/fbac4acd3677501706e3



Appendix 24: Maximum-Likelihood SNP-based phylogenetic tree of all available at EnteroBase ST131 *fimH*27 isolates (clade B). Inferred using FastTree2 and EC958 as a reference with recombination filtering applied. Tree scale bars represent the number of substitutions per site of alignment. Tip label denotes isolate name, and colored red if they known to be part of Denmark outbreak or green if they are Australia sourced. Colored ring represent geographical source of isolates. Average of 48 and 37 SNPs from Australian isolates to outbreak, while within outbreak average 2 SNPs, and overall average is 47 SNPs (51 if outbreak excluded).