Diversity and Phylogeny of Niche Adapted *Escherichia coli*

Gabriel Astorga

100231272

A thesis submitted for the degree of Doctor of Philosophy to the University of East Anglia

University of East Anglia

Quadram Institute Bioscience

September 2025

Word count: 48,676 words

Abstract

Escherichia coli is a diverse bacterial species found in the environment, on food and within the guts of humans and animals. It can be an opportunistic pathogen, cause foodborne illness, and carry antimicrobial resistance (AMR) determinants. However, the majority of *E. coli* are *commensal*, therefore it is important to distinguish the different types of *E. coli* which pose a greater health risk to humans and animals. Furthermore, understanding the environmental and host interfaces is important for investigating the dissemination and transmission of *E. coli*. Traditional typing approaches for investigating food contamination and between-host transmission offer lower resolution than whole genome sequencing (WGS). The work within this thesis investigated the evolutionary history and diversity of *E. coli* on retail foods and within hosts co-habiting within the same household using WGS. To identify the optimal phylogenomic method for investigating different *E. coli* populations, five phylogenomic approaches were applied to a test dataset of 515 genomes. Phylonium was chosen due to computational efficiency and robustness compared to the commonly utilised core gene phylogeny and applied to datasets investigated within this thesis. There were highly diverse *E. coli* isolated from five different food commodities in terms of sequence types and AMR determinants, including within individual samples. However, there were minimal putative extra-intestinal pathogenic *E. coli* identified, and more than half of isolates were pansusceptible. A household-based cross-sectional study established that there was limited *E. coli* co-occurrence between humans and their dogs. This research suggests that while retail foods as well as and co-habiting humans and animals are all potential sources of *E. coli* and AMR, the impact on human health requires further study. These outcomes were only possible through the high resolution offered by WGS. The results have implications for public health policy, particularly for food safety and One Health surveillance.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

AWaRe	Access, Watch or Reserve
AIEC	Adherent-invasive *E. coli*
AMR	Antimicrobial resistance
APEC	Avian pathogenic *E. coli*
APHA	Animal and Plant Health Agency
BAPS	Bayesian analysis of population structure
BPW	Buffered peptone water
CARD	Comprehensive antibiotic resistance database
CFU	Colony forming units
cgA	clonal group A
CGE	Center for Genomic Epidemiology
CRERB	Clinical Research Ethical Review Board
DAEC	Diffuse adhering *E. coli*
DEC	diarrheagenic *E. coli*
DEFRA	Department for Environment, Food & Rural Affairs
DNA	Deoxyribonucleic acid
EAEC	Enteroaggregative *E. coli*
ECOR	*E. coli* Reference Collection
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EMB	Eosin methylene blue
ETEC	Enterotoxigenic *E. coli*
EU	European Union
EUSPAUR	English Surveillance Programme for Antimicrobial Utilisation and Resistance
ExPEC	Extraintestinal Pathogenic *E. coli*
FSA	Food Standards Agency
GWAS	Genome Wide Association Study
HGT	Horizontal gene transfer
HH	Household
HKY	Hasegawa, Kishino, and Yano
HPC	High-Performance Computing
HUS	Haemolytic uremic syndrome
InPEC	Intestinal pathogenic *E. coli*

ISO	International Organization for Standardization
KF	Kuhner-Felsenstein measure
LMM	Linear mixed model
MDR	Multidrug resistance/resistant
MGE	Mobile genetic elements
MLST	Multilocus sequence typing
MPN	Most probable number
NCBI	National Center for Biotechnology Information
NMEC	Neonatal meningitis *E. coli*
PATH-SAFE	Pathogen Surveillance in Agriculture, Food and Environment
PBS	Phosphate buffered saline
PFGE	Pulse-field gel electrophoresis
PHE	Public Health England
POGs	Pangenome orthologous groups
RF	Robinson-Foulds distance
RNA	Ribonucleic acid
RTE	Ready-to-eat
SEPEC	Sepsis-associated *E. coli*
SNP	Single nucleotide polymorphism
SRA	Sequence read archive
ST	Sequence type
STEC	Shiga toxin-producing *E. coli*
TBX	Tryptone Bile X-Glucuronide
UK	United Kingdom
UKHSA	UK Health Security Agency
UPEC	Uropathogenic *E. coli*
UPGMA	Unweighted pair group method with arithmetic mean
UTI	Urinary tract infection
VARSS	UK-Veterinary Antibiotic Resistance Sales and Surveillance Report
VBNC	Viable but non culturable state
VFDB	Virulence factor database
VMD	Veterinary Medicines Directorate
VTEC	Verocytotoxin-producing *E. coli*
WGS	Whole genome sequencing
WHO	World Health Organization
wRF	Weighted Robinson-Foulds distance
XDR	Extensively drug resistant

Acknowledgements

To Professor Alison E. Mather, whose guidance and patience has been endless throughout this PhD. You've seen me at my lowest (thanks for calming me down during my panic attacks) and you've helped me to become a better scientist. I am grateful that you were my supervisor, my mentor, and a friend when I needed it. Thank you.

"It's a marathon, not a sprint"

To Professor John Wain, who was always pushing me to my full potential and making me think outside the box. Discussions with you are always insightful. You once told me scientists need to be creative, and I wholeheartedly agree. I'm incredibly lucky to have had you guiding me. Thank you.

"You were very rewarding"

To Dr Leonardo de Oliviera Martins, whose expertise and knowledge opened my eyes to the beauty of phylogenomics. I am grateful to have been your student, and I am excited to see the generations of students you will continue to inspire as a lecturer. You made phylogenomics fun and exciting. Thank you.

"That tree is beautiful," ***"Is it?"***

To Dr Samuel J. Bloomfield, one of the most reliable and patient people I know. You are a fantastic mentor and friend. Thank you for always helping me refine my work, teaching me constantly, and calming me down whenever I panic. I say this all the time but thank you.

"Samuel…" ***"*sigh* Yes…?"***

To Dr Dipali Singh, who kindness, patience, and knowledge inspired me continually throughout my PhD. You were always helping me be a better scientist. I remember when we had our first meeting, and you explained mathematical matrices – I was in awe. Thank you.

"If you ever need someone to talk to, we're here"

To Dr Siân Frosini, who made me feel at home during my placement at the Royal Veterinary College. You opened my eyes to the wonderful world of veterinary microbiology and taught me so much in a short amount of time. Your kindness, knowledge and skills inspire me. (I also blame this placement for my caffeine addiction).

"You absolutely are a lab person too in my eyes"

To Future Dr Oleksii Omelchenko and Future Dr Ho Yu Liu, Dr Agata Dziegiel, Dr Min Yap, Dr Samuel Bloomfield, Dr Cailean Carter, Sam Mellor, Marnie Barnham, my PhD ride or dies. I am so lucky to have made such amazing friends here at the Quadram. Thank you for the yaps, mental health walks, cries, laughter, and emotional support.

Thank you to my lab group, current and past: Dr Agata Dziegiel, Dr Min Yap, Dr Cailean Carter, Sam Mellor, Will Ovenden, Raphaëlle Palau, Dr Roxana Zamudio-Zea, Dr Ricardo Castellanos, and Dr Carol Davies-Sala. As well as Dr George Savva, Dr Sumeet Tiwari, David Baker, the core sequencing team, and the core laboratory support team.

Thank you to the Biotechnology and Biological Sciences Research Council, Norwich Research Park Doctoral Training Programme, University of East Anglia, Food Standards Agency.

And finally, thank you to my family. To my mum, Kate, Martyn and Theo. You pushed me to survive and thrive during my PhD. You continually make me a better scientist and a better person. No matter the situation, you are always there. I love you.

This PhD thesis is dedicated to my father

Mr Herminio (Pogs) Astorga

Dad, I miss you always. This one is for you.

Chapter 1 : General Introduction

Humans constantly interact with their environment and other organisms. The health of humans, animals and their environments have an interdependent link which requires collaboration across disciplines to tackle global health issues (World Health Organization, 2025). The One Health approach is an integrated approach that balances and optimises the health of humans, animals and the environment by utilising different disciplines and communities to solve health challenges (World Health Organization, 2025). This One Health approach addresses health challenges that include but are not limited to: infectious disease and antimicrobial resistance (AMR). This thesis explores the diversity and phylogeny of *Escherichia coli*, which can be found in humans, animals and the environment. In this thesis, I will explore various phylogenomic methods for investigating the genomic diversity of the *E. coli*, apply the most suitable methods to study the *E. coli* present on extra-intestinal environments such as retail foods and the *E. coli* present within human and companion animal hosts, and explore the public health implications of these findings.

1.1 *Escherichia coli*

Escherichia coli is a bacterial species that belongs to the Pseudomonadota phylum, Gammaproteobacteria class, *Enterobacteriaceae* family, and *Escherichia* genus (Yu *et al.*, 2021, Foster-Nyarko and Pallen, 2022). *E. coli* is characterised as Gram-negative (Saida *et al.*, 1998) and rod-shaped, approximately 0.5µm in diameter and 1.5µm in length (Shiomi *et al.*, 2009). The bacterium is a facultative anaerobe, meaning it can grow in both aerobic and anaerobic environments (Tenaillon *et al.*, 2010). *E. coli* is a common commensal organism that can be found within the guts of humans and animals without causing immediate harm (Tenaillon *et al.*, 2010).

The primary environment where *E. coli* resides is suggested to be the guts of humans and animals, which can then be spread to secondary extra-intestinal environments such as water and soil (Kay and Fricker, 1997). As such, *E. coli* is commonly used as a microbiological indicator in defining standards for public health and environmental control (Kay and Fricker, 1997). Within the UK, *E. coli* is used as a faecal contamination indicator on foodstuffs (European Commission, 2005) and for bathing water quality (Department for Environment Food and Rural

Affairs, 2025). As *E. coli* is typically found in faeces, the proposed route for transmission between hosts is through the faecal-oral route (Gangarosa, 1978).

1.1.1 Human infection and disease

In addition to commensal *E. coli*, the bacterium can be an opportunistic pathogen, where if the organism were to become present in a vulnerable extra-intestinal space within a vulnerable host, it may cause infection. Examples of this are bloodstream infections (UK Health Security Agency, 2025a) and urinary tract infections (Public Health England, 2007). Unlike commensal and opportunistic *E. coli*, pathogenic *E. coli* can result in disease within the host, particularly gastrointestinal disease.

When food contaminated with certain pathogenic types of *E. coli* is consumed, it can cause gastrointestinal disease shortly thereafter. These pathogenic types of *E. coli* can be grouped into pathotypes (Geurtsen *et al.*, 2022, Pokharel *et al.*, 2023) and historical classifications have relied on clinical presentations in humans, which have in turn, led to imprecise terminology and estimates (Geurtsen *et al.*, 2022). These pathotypes of *E. coli* are approximately classified into different groups: Intestinal pathogenic *E. coli* (InPEC)/diarrheagenic *E. coli* (DEC), and Extraintestinal Pathogenic *E. coli* (ExPEC) (Geurtsen *et al.*, 2022, Pokharel *et al.*, 2023) (Table 1.1). A potential source of ExPEC is the host gut itself, as faecal samples from human hosts have been shown to harbour *E. coli* that can cause extraintestinal infection (Ludden *et al.*, 2021, Nielsen *et al.*, 2014). There is also the *Shigella* species, which are genetically similar to *E. coli* (Escobar-Páramo *et al.*, 2003). Despite genetic similarity, *Shigella* and *E. coli* are often separated for epidemiological and clinical classification purposes (Chattaway *et al.*, 2017).

Table 1.1: List of the 11 pathotypes of *E. coli* with clinical presentations (Adapted from (Geurtsen *et al.*, 2022))

Niche	Pathotype	Clinical presentation
Intestinal	Enteropathogenic *E. coli* (EPEC)	Diarrhoea, vomiting, fever, dehydration

Table 1.1: List of the 11 pathotypes of *E. coli* with clinical presentations (Adapted from (Geurtsen *et al.*, 2022))

Niche	Pathotype	Clinical presentation
	Shiga toxin-producing *E. coli* (STEC); Verocytotoxin-producing *E. coli* (VTEC); Enterohemorrhagic *E. coli* (EHEC)	Mild to bloody diarrhoea, fever, abdominal cramping, vomiting, haemolytic uremic syndrome (HUS)
	Enterotoxigenic *E. coli* (ETEC)	Mild to severe watery diarrhoea
	Enteroinvasive *E. coli* (EIEC)	Profuse diarrhoea, fever, haemolytic uremic syndrome
	Enteroaggregative *E. coli* (EAEC)	Watery diarrhoea, haemolytic uremic syndrome
	Diffuse adhering *E. coli* (DAEC)	Watery diarrhoea
	Adherent-invasive *E. coli* (AIEC)	Pathogenesis of inflammatory bowel disease
Extraintestinal	Uropathogenic *E. coli* (UPEC)	Urinary tract infections
	Neonatal meningitis *E. coli* (NMEC)	Fatal meningitis in newborns
	Avian pathogenic *E. coli* (APEC)	Colibacillosis in poultry, porcine and bovine animals
	Sepsis-associated *E. coli* (SEPEC)	Bacteraemia and sepsis

Despite pathogenic *E. coli* comprising a minority of the wider population of *E. coli*, there is a bias towards focusing on these pathotypes due to the more immediate clinical relevance and public safety aspect. As of March 26th, 2025, Enterobase, an online resource and database for enteric bacteria, contained 365,787 *Escherichia*/*Shigella* genomes (Zhou *et al.*, 2020). Although databases containing genomes may have incorrect or missing metadata, of these 365,787, 8,659 were designated as "Pathogen" or "Pathogenic" status, comprising less than 2.4% of the *E. coli* database. Furthermore, in PubMed, using the search terms [*Escherichia coli*] and [Whole genome sequencing] generated 3,145 results (National Center for Biotechnology Information (NCBI), 1988). When refining the search on PubMed to include [pathogen], this led to 50.0% (1,573 results) relating to pathogenic *E. coli*. Focussing primarily on pathogenic *E. coli* will lead to underestimating the diversity within the species.

1.2 *E. coli* epidemiology, identification and typing

Due to the potential for gastrointestinal disease, there was a historical necessity to identify and differentiate the more pathogenic types of *E. coli*, as well as identify how they were spreading so preventative measures could be implemented (Robins-Browne, 1987). In the 1950s, *E. coli* surface antigens were used to identify *E. coli* outbreak strains, with the method being termed serotyping (Kauffmann, 1947). This method consisted of using antisera containing antibodies to detect and characterise specific O- polysaccharide, H- flagellar and K- capsular antigens. Currently, there are 197 different *E. coli* O- antigen serotypes and 53 H-antigen serotypes (Liu *et al.*, 2020). Serotyping is still used in *E. coli* outbreak investigations today, highlighting specific *E. coli* pathotypes (Chan *et al.*, 2023). Serotyping has been a useful approach for studying outbreaks; however, this approach has lower discriminatory power compared to other methods as this only looks at a maximum of three surface antigens compared to molecular typing approaches which can discriminate at higher resolutions.

1.2.1 Molecular typing approaches - multilocus sequence typing

Typing *E. coli* has progressed from antigen phenotyping, to the use of enzymes (Whittam *et al.*, 1983, Selander *et al.*, 1986), ribosomal RNA genes (Martin *et al.*, 1996), and DNA fragment-based approaches (Gautom, 1997). Currently, multilocus sequence typing (MLST) is one of the most common *E. coli* typing approaches, performed by examining the sequence of multiple housekeeping genes. This method can utilise various MLST schemes, for example, two-locus *fumC* and *fimH* (CH) typing, and four-locus (*adk, purA, gyrB* and *recA*) typing (Ahmed *et al.*, 2016). One of the most common *E. coli* typing schemes is the Achtman seven gene scheme which originally used internal fragments from the *adk* (adenosine kinase), *fumC* (fumarse C), *gyrB* (DNA gyrase subunit B)*, icd* (isocitrate dehydrogenase), *mdh* (malate dehydrogenase)*, purA* (adenylosuccinate synthetase) and *recA* house-keeping genes for PCR amplification (Wirth *et al.*, 2006, Clermont *et al.*, 2015). Internal fragments of each gene are used (approximately 450 – 550 bp) and for each house-keeping gene, a single nucleotide polymorphism (SNP) within any of these gene would lead to a different allele, and therefore a different ST allocation (Wirth *et al.*, 2006).

The number of STs have increased over the years and a more recent 2021 paper analysing 10,667 genomes identified 1,155 unique STs (Abram *et al.*, 2021). The first mention of an *E. coli* clonal complex was in a paper by Zhang *et al.,* (2000) which described a clonal *E. coli* group based on the H serogroup (Zhang *et al.*, 2000). In 2001, clonal group A (cgA) was reported as a group of *E. coli* found in urinary tract infections with similar DNA fingerprints, virulence factor profiles, antimicrobial resistance, and O antigen groups (Manges *et al.*, 2001). The first use of the current MLST based clonal complexes, however, was in a 2005 study by Tartof *et al.,* which classified a group of uropathogenic and bacteraemic *E. coli* into a single ST69 complex (Tartof *et al.*, 2005). Since then, 54 ST clonal complexes have been documented (Zhou *et al.*, 2020).

The combination of seven house-keeping neutral markers in MLST has limitations for differentiating *E. coli* at higher resolutions. It is also not uncommon for MLST classifications to be combined with other subtyping analyses, like serotyping or *fimH* typing, which is types *fimH* alleles (Roer *et al.*, 2017). This has been useful in the case where the specific *E. coli* have needed sub-ST level differentiation for epidemiological tracking (Giufre *et al.*, 2014). An example would be ST131, where different variations in the O and H antigens have been used to identify sub-lineages that pose different levels of risk if there is an infection (Decano and Downing, 2019, Pitout *et al.*, 2022). Overall, MLSTs have provided insight into *E. coli* diversity and population structures. With whole genome sequencing (WGS), higher resolution insight into *E. coli* diversity can be obtained by sequencing the entire genome, rather than just specific genes.

1.2.2 *E. coli* – Genomics

A genome comprises all an organism's genetic information (Goldman and Landweber, 2016). Deoxyribonucleic acid (DNA) is comprised of four bases, which is ordered in an organised sequence (Watson and Crick, 1953). WGS is the process where the base order of the DNA sequence is called, and the data are collected, which can be analysed by downstream bioinformatic approaches (Heather and Chain, 2016). Furthermore, advancements in next generation and third generation sequencing can now generate large amounts of accurate genomic data rapidly, allowing whole populations of organisms to be analysed (Lu *et al.*, 2025). Currently WGS is the highest resolution typing approach (Figure 1.1).

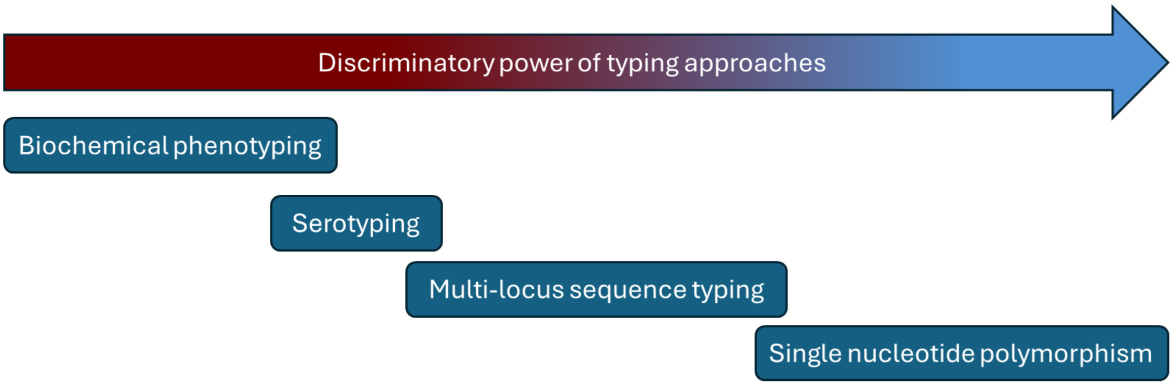


Figure 1.1: A non-exhaustive list of different typing approaches used classify *E. coli*. The arrow indicates increasing discriminatory power.

E. coli has a single circular chromosome located within the central nucleoid region of the cell. The length of the *E. coli* genome is variable and when *E. coli* populations have been investigated (including *Shigella* as a subspecies of *E. coli*), the *E. coli* genome length has been found to range between 4.56 Mbp to 5.70 Mbp (Lukjancenko *et al.*, 2010). The reason for this are large sections of DNA that can be present in one genome, which can be absent in another different genome, termed accessory segments or genes (Rode *et al.*, 1999).

The pangenome is the set of all genes in all the genomes of a group of organisms being investigated (Bosi *et al.*, 2015) and is comprised of the core and accessory genome (Rasko *et al.*, 2008). *E. coli* has an open pangenome, where new genes are generally added when new *E. coli* genomes are included in the dataset (Touchon *et al.*, 2009). Additional *E. coli* added to the collection are likely to have novel genes that are not present in the previous compared set (Rasko *et al.*, 2008). The core genome is defined as the set of conserved homologous genes found in a specified percentage of a given set of genomes (Rasko *et al.*, 2008).

The size of the core genome is dependent on the total number of investigated isolates and the specified percentage of isolates containing those homologous genes. Therefore the *E. coli* core genome can vary between studies, and what may be defined as present in all organisms in one dataset, may not necessarily be true for another *E. coli* collection. For example, a study looking at 53 *E. coli* genomes found that 11% of the genome was conserved, dropping down to 6% when *Shigella* genomes were included (Lukjancenko *et al.*, 2010). In contrast, a study looking at 4,071 *E. coli* isolates found that 14% of the pangenome was comprised of core genes (Decano and Downing, 2019). This represents genes that are essential for species survival, central metabolism, replication, translation, and transcription (Medini *et al.*, 2005, Rasko *et al.*, 2008).

1.2.3 *E. coli* mobile genetic elements

The accessory genome refers to all the non-conserved genes in the pangenome (Segerman, 2012). The diversity of accessory genes in *E. coli* is due to gene gain from horizontal gene transfer (HGT) and gene loss, which is hypothesised to be driven by environmental adaptation (McInerney *et al.*, 2017). HGT refers to the acquisition of external DNA from one organism to another, which is then incorporated into the recipient organism most often via the process of recombination (Lawrence and Retchless, 2009). Genes gained from HGT can be beneficial for utilising different substrates, gaining stress tolerance and facilitating energy efficient growth (Karcagi *et al.*, 2016). Not all genes gained via HGT are beneficial to growth and adaptation, and therefore there may be a selective process for advantageous horizontally transferred genes and linked neutral genes (Vogan and Higgs, 2011).

E. coli mobile genomic elements (MGEs) include transposons, phage and phage-like structures, and plasmids (Jackson *et al.*, 2011). Integrons are non-mobile genetic elements that are linked to mobile genetic elements (MGEs) and can act as a vector for genes important for adaptation, such as AMR genes (Sabate and Prats, 2002, Domingues *et al.*, 2012). Similarly, plasmids are extrachromosomal circular strands of DNA that can act as vectors for accessory genes, including AMR genes.

1.2.4 Sentinel organism for antimicrobial resistance

Antimicrobials are compounds that stop the growth or kill a range of microorganisms (UK Health Security Agency, 2024). Microorganisms can adapt to selection pressures like antimicrobials and develop mechanisms to resist the effects of these drugs (Christaki *et al.*, 2020). Antimicrobial resistance (AMR) is a public health threat as it can make current treatments less effective, prolong the duration of infections (GOV UK, 2024). Furthermore, a delay in appropriate antimicrobial treatments can lead to increased mortality rate (Andersson *et al.*, 2019).

AMR is one of the top global health threats, predicted to be associated with an estimated 4.95 million deaths, which include an estimated 1.27 million deaths attributed to AMR bacteria (Murray *et al.*, 2022). In 2023, antibiotic-resistant infections within the UK have risen compared to 2019; *E. coli* was the greatest contributor, responsible for over 65% of antibiotic-resistant bacteraemia cases between 2019-2025 (UK Health Security Agency, 2024). In another study of 14,548 *E. coli* bloodstream infection cases, AMR was associated with increased 90-day mortality amongst patients (Daneman *et al.*, 2023).

Currently, a coordinated antimicrobial stewardship approach for promoting responsible selection, usage, dosage and duration of antimicrobials for optimal clinical outcomes is led by the World Health Organization (World Health Organization, 2023). An antimicrobial classification system called the ‘AWaRe (Access, Watch or Reserve)’ list created by the WHO was recently adopted by the UK Health Security Agency to support antimicrobial stewardship at the local and national level (UK Health Security Agency, 2025b). This list has helped inform UK government reports and is expanded on in **section 1.4.1** of this thesis. Within these reports, *E. coli* is reported as it is a commonly used sentinel organism for AMR.

E. coli can acquire resistance to different antibiotics through different mechanisms (see Section 1.2.3). The health risk is amplified by the potential for multidrug resistance (MDR), which is proposed in *Enterobacteriaceae* to be an isolate that is resistant to at least one drug in at least three or more antimicrobial classes (Table 1.2) (Magiorakos *et al.*, 2012). Concerningly, *E. coli* can potentially become extensively drug resistant (XDR) (Guy *et al.*, 2023), characterised by resistance to nearly all antibiotic categories.

Table 1.2: A non-exhaustive list of antibiotics and the drug target

Antibiotic	**Target**
Penicillin	Cell wall synthesis
Cephalosporins	
Carbapenems	
Monobactams	
Glycopeptides	
Phosphonic acids	
Polymyxins	Cell membrane
Aminoglycosides	30S ribosome
Tetracyclines	
Oxazolidinones	50S ribosome
Streptogramins	
Macrolides	
Lincosamides	
Phenicols	
Quinolones/	DNA topoisomerase

Table 1.2: A non-exhaustive list of antibiotics and the drug target

Antibiotic		Target
Fluoroquinolones		
Sulphonamides		Folic acid synthesis inhibitors
Trimethoprim	Dihydrofolate reductase (DHFR) in folic acid synthesis inhibitor	

AMR risk is not just limited to pathogenic *E. coli* and opportunistic infections. Commensal *E. coli* within the guts of humans and animals have been identified as potential reservoirs for AMR. For example, AMR genes, MDR and putative plasmid-carriage of AMR have been identified in healthy humans (Tawfick *et al.*, 2022, Bailey *et al.*, 2010). Commensal *E. coli* maintains an important role in the dissemination of AMR, as *E. coli* within the gut can spread AMR genes to other *E. coli* and other bacteria through mechanisms like HGT.

1.2.5 *E. coli* population genomics

Population genomics aims to understand the genomic variation across different populations of an organism by comparing their genomic information (Xu, 2006). Like other bacteria, *E. coli* reproduces by binary fission, an asexual process generating near genetically indistinguishable clones from their parent with the exception of random mutations (Zhi *et al.*, 2019). It is proposed that genetic variation in a bacterial species like *E. coli* is due to established evolutionary mechanisms like genetic drift, gene flow, natural selection, and random mutations (Gordo *et al.*, 2014).

Early *E. coli* population structure studies suggested there was little evidence of frequent recombination and that the *E. coli* population structure consisted of independently evolving clones (Sears *et al.*, 1950). As such, genetic variation within these *E. coli* clonal lineages were thought to be due to the accumulation of vertically transmitted mutations. However the genomic diversity between *E. coli* lineages is now identified to be caused more by recombination than accumulation of vertical mutations (Walk *et al.*, 2007). This is suggested by linkage disequilibrium, which is the association of alleles in dissimilar positions on a chromosome in a non-random fashion (Slatkin, 2008). Genetic elements acquired through HGT can be integrated into the bacterial genome through recombination, which would break associations between alleles. Recombination can be a potential barrier for investigating the evolutionary history of an organism. Areas of the genome acquired from HGT, and recombination may have different evolutionary histories, and when compared to the true species evolutionary history, may lead to incongruence when determining the evolutionary history.

1.3 Phylogenomics

Phylogeny aims to reconstruct the evolutionary history of a group of organisms based on shared characteristic traits like physical morphology, amino acids, or genetic sequence data (Lee and Palci, 2015, Schmidt, 1995, Lees *et al.*, 2018b, Sakoparnig *et al.*, 2021). The basis of phylogeny stems from homology, which is the similarity of a compared organism, object or concept often derived from a shared origin.

Phylogenetics is the study of the evolutionary history of a group of organisms using the genetic information within organisms. Within phylogenetics, the term homologous genes or proteins are used, which are DNA or amino acid sequences from distinct lineages that have a common ancestor. These shared genes may arise in different lineages via two ways: a) speciation events, when a subset of a population become separated from other members of the original population and develop unique characteristics, creating orthologous genes (orthologs); b) gene duplication events, where genes are copied within the genome leading to multiple-copy genes termed paralogs. Additionally, acquisition of similar genes can arise through HGT (Gevers *et al.*, 2004). This process may allow previously external homologous genes, termed xenologs, to become integrated into the host genome or acquired through plasmids (Price *et al.*, 2008).

Phylogenomics builds on this concept and explores evolutionary and genetic relationships by using the whole genome and orthologous loci, models of evolution and statistical inference (Young and Gillung, 2020, Nascimento *et al.*, 2017). Species phylogenies are typically denoted via a tree, where terminal ‘tip’ nodes represent the operational taxonomic units, and the internal nodes denote a theoretical ancestor. Lineages of an organism can be defined as descendants from this common theoretical ancestor and can be grouped into clades (Figure 1.2). Species trees can also be referred to as cladograms, where the species-ancestor relationship is shown, or dendrograms where the taxonomic relationship between isolates is highlighted and the phylogenetic relationship is hidden (Brower, 2016). Phylogenies are not restricted to just groups of organisms but can also be used to understand the evolution of genomes and families of genes. Due to the advancements in sequencing technologies and costs, whole genome scale data can now be used to evaluate the evolutionary history of a group of organisms. Phylogenomics builds on phylogenetics and utilises large portions of the genome or the whole genome, whereas phylogenetics analyses the sequences of single genes or a small number of genes or trait data (Delsuc *et al.*, 2005).

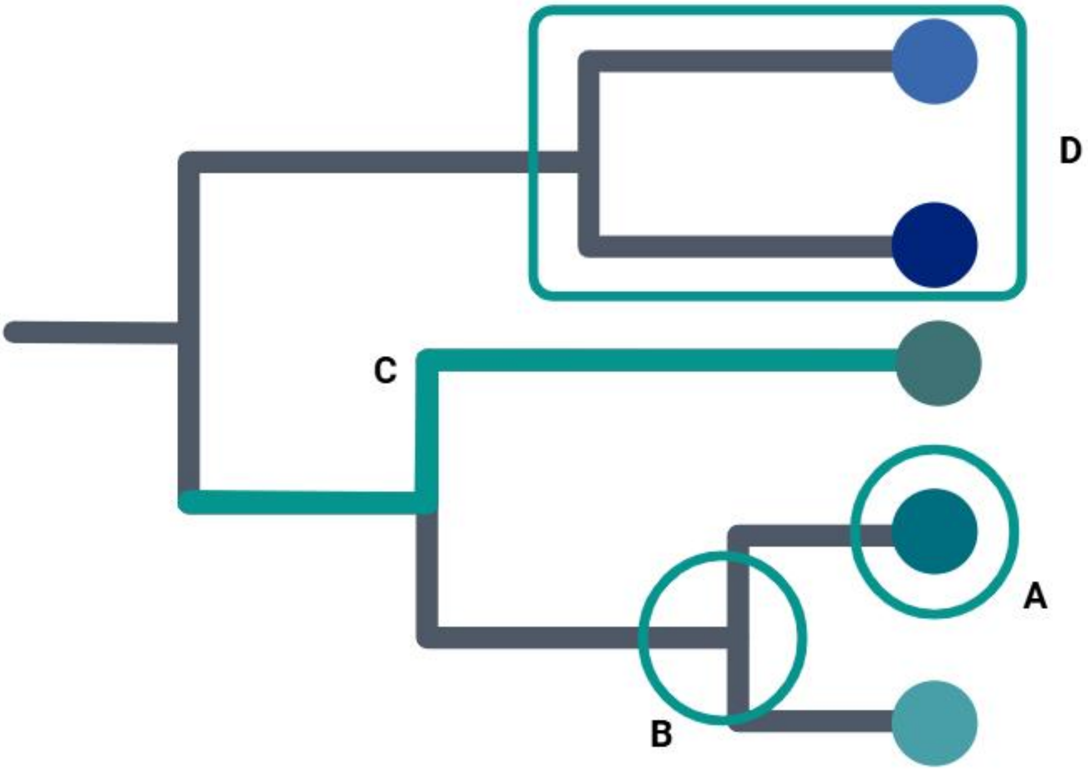


Figure 1.2: A theoretical species phylogenetic tree with the structure highlighted and explained: A) Tip (operational taxonomic unit (OTU)), B) Internal node (theoretical ancestor), C) Lineage (connection between

ancestor to the OTU), and D) Clade (a group of organisms that descend from a common ancestor) (Created with BioRender.com).

1.3.1 *E. coli* phylogroups

There is a broad repeatable phylogenetic population structure that has been identified for *E. coli*. The *E. coli* Reference Collection (ECOR) was created as a set of reference isolates from differing hosts and locations to start exploring this diversity (Ochman and Selander, 1984). The Clermont triplex PCR method, built on the ECOR set, was used to classify *E. coli* into phylogroups, which were subsequently developed in further studies to produce a final set of A, B1, B2, C, D, E, F and G phylogroups (Clermont *et al.*, 2000, Clermont *et al.*, 2013, Clermont *et al.*, 2019). Specific *E. coli* lineages did not fall easily into these phylogroup categories but were still classified as *E. coli* - *E. coli* cryptic clades I-VI are lineages of *E. coli* that are taxonomically and phylogenetically distant from the other phylogroups but have the same typical biochemical markers (Walk *et al.*, 2009, Gangiredla *et al.*, 2018). Cryptic clades III and IV have been classified as a separate species, *Escherichia ruysiae* (van der Putten et al., 2021), and cryptic clade V was reclassified as *Escherichia marmotae* (Liu *et al.*, 2015). The *E. coli* phylogroups are a current way of phylogenetically classifying *E. coli* into grouped lineages. However, this current phylogenetic typing scheme is based on eight target genes, which do not take phenotypic characteristics or the accessory genome into account.

These phylogroups have been useful for examining the evolutionary relationship between *E. coli* genomes. The results from a MASH-based analysis of 10,667 *E. coli* genomes were in line with previous results by Gonzalez-Alba *et al.,* suggesting phylogroup D was most closely related to the estimated origin of *E. coli*, then the E, B1 and A phylogroups as another lineage that has differentiated, and finally the F, G and B2 lineages (Abram *et al.*, 2021, Gonzalez-Alba *et al.*, 2019). Furthermore, the variation in *E. coli* genome sizes is directly correlated with phylogroup classification, with the most recently diverged phylogroup F having the largest chromosomal size (Gonzalez-Alba *et al.*, 2019).

Broadly speaking, each phylogroup has been associated with distinct ecological niches, although there has been some overlap. *E. coli* from phylogroup A and B1 are often associated with a commensal nature, being the predominant phylogroup within healthy humans (Hazen *et al.*, 2023, Duriez *et al.*, 2001, Lescat *et al.*, 2013). Phylogroup B1 has also been found to be predominant in wild animals, livestock (Lescat *et al.*, 2013, Coura *et al.*, 2015, Muloi *et al.*, 2022) and environmental aquatic samples (Ratajczak *et al.*, 2010, Touchon *et al.*, 2020). Phylogroups B2 and D have been associated with clinical infections and as extraintestinal pathogens (Russo and Johnson, 2000, Picard *et al.*, 1999). Phylogroup E has been found within a range of human, bovine and avian hosts, and have been found to be genetically diverse (Clermont *et al.*, 2021). Phylogroup F has been primarily associated with poultry and the extra intestinal pathogenic lineages have been termed Avian Pathogenic *E. coli* (APEC) (Zhuge *et al.*, 2021). Phylogroup G has been designated an intermediate group of phylogroups F and B2, displaying similar extra intestinal virulence and associations with poultry as their host (Clermont *et al.*, 2019). It was suggested by Gonzalez-Alba *et al.,* that phylogroup-specific ecological adaptations may be due to the gain of metabolic flexibility for the E, B1 and A lineages, and loss of environmental-adaptive mechanisms in the F, G2 and B2 lineages (Gonzalez-Alba *et al.*, 2019).

E. coli from the same phylogroup and that have the same ST can often display different accessory genomes. For example, a 2016 study by McNally *et al.,* found that ST131 can be further split into three phylogenetic clades, which could be further split by distinct clusters based on non-conserved accessory genes. These accessory genome clusters were found to have unique combinations of genes that circulate within subtypes of ST131 via HGT (McNally *et al.*, 2016).

E. coli population structures within specific niches can be shaped by host and environmental factors. These population structures may not be fully clonal and monophyletic but can still cluster by phylogroup (Tenaillon *et al.*, 2010). The *E. coli* clonal group concept is not yet fully defined as phylogroups based on lineage differentiation still contain multiple lineages that show genetic variation. Population genetics studies have explored host-*E. coli* relationships (Duriez *et al.*, 2001, Muloi *et al.*, 2022), revealing that certain phylogroup and population structures are influenced by certain hosts, host interactions, geographical locations, and temperatures (Muloi *et al.*, 2022). In a 2020 paper by Touchon *et al., E. coli* phylogroups were found to be associated with genome size and isolation sources. The isolation source was also found to influence the number of MGEs, suggesting that the niche the *E. coli* are isolated from may influence genomic diversification (Touchon *et al.*, 2020).

1.4. Host origins and contamination sources

The gut, where *E. coli* typically resides, presents as a highly competitive environment, with multiple microorganisms competing for the same nutrients (Biggs *et al.*, 2017, Medlock *et al.*, 2018). Generally, *E. coli* can access and utilise nutrients present in the mucus like iron and various sugars (Cohen *et al.*, 1991, Leatham *et al.*, 2005). However, specific lineages of the species can differ in metabolic capabilities, which is evident within *in vivo* mice models, when pathogenic *E. coli* O157:H7 was identified to utilise galactose, hexuronates, mannose, and ribose, whereas *E. coli* MG1655 was not (Fabich *et al.*, 2008). Additionally, *E. coli* exist with other microbes and are able to work synergistically with differing *E. coli* lineages and other bacterial species to survive (e.g., forming biofilms) in laboratory settings (Reisner *et al.*, 2006, Kaushik *et al.*, 2025, Sung *et al.*, 2025).

1.4.1 *E. coli* in humans

Early *E. coli* population studies within humans described how individual human hosts can harbour multiple different *E. coli* clones (Selander and Levin, 1980), and argued that the human gut contained transient and resident *E. coli* strains using lower resolution serotyping or electrophoresis approaches (Caugant *et al.*, 1981, Sears and Brownlee, 1952). Caugant *et al.* also identified that the diversity of *E. coli* types within a host could drastically vary from human to human and argued in this early work that the genetic diversity is a consequence of the acquisition of evolutionary divergent *E. coli*. Evidence for this hypothesis was identified using higher resolution WGS approaches, where healthy humans harboured multiple different genotypes of *E. coli*; Foster-Nyarko *et al*. sequenced up to five *E. coli* colonies and identified an average 2.7 different genotypes within a host (Foster-Nyarko *et al.*, 2021). The presence of multiple different lineages of *E. coli* within a host is suggested to be more influenced by immigration events, whereas within-host evolution only plays a minor role in the diversity observed (Foster-Nyarko *et al.*, 2021).

Within healthy individuals, the rate at which different *E. coli* phylogroups are replaced can vary; in agreement with the idea that there are transient and resident lineages, whole genome analysis from a 20-year longitudinal gut study suggested that there are two rates of turnover – rapid, whereby specific clones are present for a week, and slow, in which carriage can last several years (Condamine *et al.*, 2025). When categorised into phylogroups, Condamine *et al.,* (2025) also identified that when phylogroup B2 was predominant, other phylogroups tended not to co-occur, and when phylogroup A was predominant, other phylogroups could co-exist. This is contrasted by other work by Morel-Journel *et al.,* which indicates that if there are more diverse *E. coli* communities, the rates of new *E. coli* colonisation increase, regardless of the specific phylogroups present (Morel-Journel *et al.*, 2025). Morel-Journel *et al.* (2025) also suggested that longer phylogroup residence within the gut has a colonisation trade-off, whereby *E. coli* that are better at maintaining residence have weaker initial colonisation capacity compared to more transient lineages.

The diversity observed within the human gut is suggested to originate from immigration events, whereby *E. coli* is introduced through other hosts and the environment (Dixit *et al.*, 2018, Foster-Nyarko *et al.*, 2021). As *E. coli* is typically a gastrointestinal bacterium, there are multiple proposed routes of exposure that can lead to potential transmission to humans: a) environmental transfer from faecal contamination on food and water (Kay and Fricker, 1997, Castro-Rosas *et al.*, 2012); b) transfer from other hosts like humans or animals (García *et al.*, 2010, Johnson *et al.*, 2016).

1.4.2 *E. coli* in animals

1.4.2.1 *E. coli* within companion animals

A companion animal is defined in this thesis as any domesticated animal living permanently within a community and kept by humans for purposes such as companionship, work, or psychological support (Pongrácz and Dobos, 2023). Companion pets are highlighted as one of many potential animal reservoirs for zoonotic disease by the UK government (Animal and Plant Health Agency, 2025). Companion animals are a part of many UK homes and in 2018, there was an estimated 11.1 million cats and 8.9 million dogs within UK households (Wensley *et al.*, 2021). The shared environment and lifestyle of the owner can affect the daily habits of the pet and what the companion pets are exposed to, and vice versa (Cunningham-Smith and Emery, 2020).

Companion dogs and cats have been identified to harbour a range of different *E. coli* lineages on different parts of their body; *E. coli* isolated from the nasal, oral, rectal, abdomen and hindquarters of a group of 155 dogs and 121 cats could harbour AMR *E. coli* from different phylogenetic clades (Davis *et al.*, 2011). Within healthy dog populations, phylogroups A, B1, B2, D, E and F have been recovered (Bourne *et al.*, 2019, Schmidt *et al.*, 2015). Current WGS studies evaluating *E. coli* populations in healthy dogs have generally focused on AMR *E. coli* (Saria *et al.*, 2025, Sealey *et al.*, 2023, Boehmer *et al.*, 2018).

Specific sub-lineages of select STs have been investigated in more detail and assessed for host association; clades of ST73 have been found to be found specifically in companion pets and not humans (Kidsley *et al.*, 2020a). This contrasts with ST131, which is consistently highlighted as present within humans, and can also be present within companion animals. Specific clades of ST131 pose different potential health risks in terms of ExPEC and AMR potential (Kidsley *et al.*, 2020b, Pomba *et al.*, 2014, Timofte *et al.*, 2014, Johnson *et al.*, 2016). ST372 *E. coli* is identified as a dominant ExPEC lineage in dogs, argued to have sub-lineages adapted to humans through the loss of the *pdu* propanediol utilization operon, suggested to be linked to differences in host species diet (Elankumuran *et al.*, 2023).

Work investigating *E. coli* within companion dogs has also focused on *E. coli* sharing within households between humans and dogs (Caugant *et al.*, 1984, Johnson *et al.*, 2001). Both healthy humans and healthy dogs can be colonised by *E. coli,* and it has been suggested that cohabitation and repeated exposure may promote *E. coli* sharing and long-term colonisation (Johnson *et al.*, 2008, Johnson and Clabots, 2006). Low resolution typing approaches by Caugant *et al.* (1984) and Johnson *et al.* (2001), and a smaller study by Johnson and Clabots (2006) and Johnson *et al.* (2008) identified that co-occurrence happened within and between species within households. Johnson *et al.* (2008) further observed that the *E. coli* co-occurrence was more prevalent within households than between households.

In terms of indirect sources, Johnson *et al.* (2008) proposed the shared-food hypothesis, whereby both hosts are exposed to an external reservoir. Food represents an exposure route for both humans and animals. Previous work studying raw meat diets for pets strongly suggests that raw meat diets may introduce AMR *E. coli* to companion dogs, leading to differences between raw meat fed dogs and non-raw meat fed dogs (Groat *et al.*, 2022, Schmidt *et al.*, 2015, Morgan *et al.*, 2024). However, this would be fundamentally different for humans as meat consumed by people would usually be cooked (retail food contamination and exposure is expanded on in **Section 1.4.5**). Additionally, the environment (e.g., household surfaces) that the hosts interact with can also be a potential interface for exposure, which has been suggested to play a role in the dissemination of MDR *E. coli* (Leite-Martins *et al.*, 2015).

The domestic setting has been highlighted by Public Health England (PHE) in coordination with key UK government agencies and public health bodies, as a way zoonotic disease can arise (Public Health England, 2016). The collection of national surveillance data for livestock, wildlife and small animals can be improved:

1) The APHA surveillance approach is based on submissions to Veterinary Investigation Centres and private surveillance providers contracted by the APHA. According to the Public Health Act 1984, notifiable diseases need to be reported (Public Health (Control of Disease) Act 1984 c.22, 1984). Bacteria are isolated from diagnostic submissions from diseased animals (i.e., clinical samples), and post-mortem samples, but not all samples are submitted to government laboratories for further analysis (Animal and Plant Health Agency, 2025).
2)) The APHA does not provide a veterinary diagnostic testing or post-mortem examination for companion animals outside of *Salmonella, Yersinia enterocolitica,* and general aerobic and anaerobic cultures (Animal and Plant Health Agency, 2024).

There are currently no AMR surveillance systems in place for healthy dogs and cats in the UK (UK Health Security Agency and Veterinary Medicines Directorate, 2023). However, as of July 2025, a four-year pilot study was announced by the

Veterinary Medicines Directorate (VMD) that will address gaps in clinical surveillance data in companion animals (Veterinary Medicines Directorate, 2025). Although sales data on veterinary antibiotics are collected and analysed (Veterinary Medicines Directorate, 2024), there is a need for comprehensive assessment of how this is reflected in AMR levels in companion animals. The potential for zoonotic sharing and health risks between dogs and humans have been well documented – for example the detection of human-associated ST131 *E. coli* containing a *bla*CTX-M-15 gene that had conjugation capability (Timofte *et al.*, 2014). Companion dogs have also been identified as a reservoir for AMR *E. coli*; for example, in a study of 73 healthy companion dogs, 63.9% (n = 46 faecal samples) had at least one *E. coli* with AMR, and 30.6% (n = 22 faecal samples) had MDR *E. coli* (Schmidt *et al.*, 2015). Interspecies sharing of AMR *E. coli* is a risk and has been identified in a small number of cases. A longitudinal study focusing on ESBL/AmpC-producing Enterobacterales within the UK and Portugal identified two instances of a *bla*CTX-M-15 (O8:H9-ST410) and a *bla*CTX-M-55/CMY-2-producing (O11:H25-ST457) *E. coli* interspecies sharing (Menezes *et al.*, 2023).

By assessing the sharing and AMR risk using commensal *E. coli* within a household setting, a greater understanding of the risk that this multi-host opportunistic pathogen can present to both humans and pets can be achieved. WGS population studies assessing sharing and AMR are a promising tool for progressing understanding of the risks to both hosts.

1.4.3 *E. coli* on retail foods

According to the World Health Organisation Technical Advisory Group for Food Safety, contaminated food resulted in an estimated 600,000,000 global illnesses in 2010 (World Health Organization, 2022, World Health Organization, 2015). In the United Kingdom, the Food Standards Agency reports an estimated 25% of the population suffer from an intestinal infectious disease annually (Tam *et al.*, 2012). As evident by the number of foodborne illnesses, current approaches for assessing food safety can be further improved. Of the four most significant bacterial pathogens on food, the only bacteria where a specific subtype of a species is specified is *E. coli* because this subtype is known to cause life-threatening gastrointestinal disease (GOV UK, 2021). Food act as a fomite - non-living vector that can aid in the dissemination of an infectious disease (Hamilton-Miller, 2005).

The food manufacturing process differs greatly between different meat products, different seafood products and different types of leafy vegetables. To produce raw chicken meats, chickens are stunned, killed, suspended in air, bled out, mechanically defeathered, and undergo multiple scalding, cutting and chilling processes (Rouger *et al.*, 2017). The wide range of pork products on the market means that the manufacturing processes can differ, thus changing potential points of contamination or niche conditions that the bacteria have to overcome (Lebret and Čandek-Potokar, 2022). Post slaughter, pig carcasses are suspended in the air, then cut into shoulder steaks, loins, sides, and legs (Keenan, 2016). The salmon fillet processing route also goes through a similar process, which includes stunning, bleeding, gutting, filleting, possible skinning, packaging and storage (Møretrø *et al.*, 2016) Prawns require different processing to retail meats and they are caught from either wild or aquaculture environments, frozen and packaged, ready for distribution or further de-shelling and/or cooking (Farmery *et al.*, 2015).

Stress to bacteria can be defined as a harmful condition that can negatively impact survival in a niche (Chung *et al.*, 2006, Begley and Hill, 2015). The constant thermal, osmotic, chemical and biological stress that bacteria are exposed to on food suggests that the microorganisms that survive have acquired or developed mechanisms to overcome these events (Chung *et al.*, 2006). These events influence the lineages present on food by acting as genetic bottlenecks, where there is a selection for certain genes and phenotypes that are advantageous to survival in the niche. The survival of *E. coli* on retail foods depends on the ability to survive in changing temperature, osmotic conditions, oxygen conditions, chemical disinfectants and low pH (Chung *et al.*, 2006). Pathogenic lineages of *E. coli* have been found to survive these extreme conditions on food processing surfaces by expressing genes that allow successful attachment, biofilm formation (Álvarez-Ordóñez *et al.*, 2013) or existence in a viable but non culturable state (VBNC) (Dinu and Bach, 2011).

The safe threshold for microbiological contamination on food is determined using enumeration methods (International Organization for Standardization, 2001, International Organization for Standardization, 2015). Detecting the amount of *E. coli* using colony forming units (CFU) on food is common practice for assessing food safety (Food Standards Agency, 2023, Health Protection Agency, 2009). The presence of *E. coli* on retail foods can indicate poor food safety practices. For example, the presence of *E. coli* on food has been associated with cross-contamination and ineffective cleaning measures (Holvoet *et al.*, 2014, Jensen *et al.*, 2015), and uncontrolled temperatures and time (Luo *et al.*, 2010). The safe contamination thresholds depend on the food commodity and whether the food is designated ready-to-eat (RTE) or not an RTE food (European Commission, 2005).

Enumeration does not take the genetic diversity of *E. coli* into account and different types of *E. coli* pose different immediate gastrointestinal risk (Geurtsen *et al.*, 2022). For example, the infectious dose from an *E. coli* O157 outbreak was estimated to be between two to nine CFU, whereas an enterotoxigenic *E. coli* outbreak had an estimated ingestion dose was 25 – 1000 CFU (Hara-Kudo and Takatori, 2011). Therefore, enumerating safe levels of *E. coli* contamination on food lacks the ability to distinguish between *E. coli* that may present a greater health risk to consumers.

Population studies looking at commensal *E. coli* found on retail raw meat products observed a wide range of *E. coli* STs (Hayashi *et al.*, 2018, Yamaji *et al.*, 2018). There have been no studies on large populations of *E. coli* STs on prawns and salmon. Additionally, other seafood related studies have focused on typing specific lineages containing certain AMR genes of interest only (Ellis-Iversen *et al.*, 2020, Roschanski *et al.*, 2017). Therefore, the common *E. coli* STs on retail seafood and whether or not certain STs are associated with this specific food niche are currently unknown. A WGS population study investigating tetracycline-resistant *E. coli* on store-bought leafy green produce by Reid *et al.* classified 120 *E. coli* genomes into 21 STs with varied accessory genomes (Reid *et al.*, 2020). These papers discuss the range of *E. coli* STs recovered, but do not explore whether these STs represent the whole population. Therefore, further work is needed to identify the variation in *E. coli* lineages within these different retail foods.

1.5 Aims and objectives

Understanding the population structure and diversity of *E. coli* in different environments can help us understand how different lineages may be adapted across a range of compartments. This is also useful for a bacterium like *E. coli*, which can be pathogenic and/or spread AMR determinants. When investigating reservoirs of *E. coli*, we can look at not just the different types of *E. coli* present, but also the accessory genes which may facilitate adaptation or distinguish *E. coli* from different sources.

Using WGS, this project aimed to address the knowledge gaps outlined throughout this chapter:

1) Assess alternative and optimal phylogenomic approaches that can be used to investigate the diversity of *E. coli* in future studies.
    - Approaches will consider other aspects of the genome outside of the core genes.
    - Will create alternative phylogenomic approaches considering individual gene histories.
    - Assess which approaches are appropriate for future use.
2) Investigate the diversity and risk of *E. coli* on retail foods using a multi-isolate non-AMR selective approach.
    - Identify whether or not current enumeration approaches are sufficient for food safety.
    - Identify and compare *E. coli* diversity within and between different food commodities, as well as within food samples.
    - Identify potential AMR risk represented by retail foods.
3) Investigate potential interspecies sharing between dogs and humans within and between the same household and therefore assessing risk to both hosts.
    - Identify the diversity of *E. coli* within dogs and humans, as well as putative sharing events within and between hosts.

- Identify any putative host-associated genomic signatures which can be further investigated for identify potential adaptive signatures.
- Identify genes of concern, such as those conferring AMR, from *E. coli* isolated from humans and dogs.

Chapter 2: Materials and Methods

Chapter contributions: Sample collection, design, processing, and *E. coli* culturing and isolation for each specific dataset is specified in each chapter. Short read whole genome sequencing was performed by the Quadram core sequencing team. Bioinformatic quality control processes were taught by Bloomfield, S.J. R analyses were taught by Zamudio-Zea R., and de Oliviera Martins, L. Bioinformatic analyses was done by Astorga, G. Work was discussed by Astorga, G., Mather, A.E., Singh, D., Wain, J., Bloomfield, S.J., and de Oliveira Martins, L.

2.1 *E. coli* datasets

Two separate *E. coli* collections were used for this thesis:

1) A collection of *E. coli* isolated from retail foods from Norwich, UK, of which a subset was used to compare different phylogenomic approaches.

   The design of the food survey and sample collection was done by Mather, A.E., Janecko, N., Bloomfield S.J., and Palau, R. The collection of the food samples, as well as growth, detection and isolation of the *E. coli* were done by Bloomfield S.J., and Palau, R. The data analysis, scripts and code were written by Astorga, G. with help from Zamudio, R., Bloomfield, S.J., and de Oliveira Martins, L. All the work was discussed with Astorga, G., Mather, A.E., Singh, D., Wain, J., Zamudio, R., Bloomfield, S.J., and de Oliveira Martins, L. The comparison of phylogenomic methods of *E. coli* from retail foods in Norfolk, United Kingdom was planned and developed by Astorga, G. and de Oliveira Martins, L.

2) A collection of *E. coli* from human and companion dog faecal samples provided by a collaboration with the Royal Veterinary College.

   The design of original survey and sample collection was done by Menezes, J., Pomba, C., Frosini, S.M., and Loeffler, A., in collaboration between the University of Lisbon and the Royal Veterinary College. The study design, growth, detection and isolation of the *E. coli* were done by Astorga, G., and Frosini, S.M. All of the scripts, code, and bioinformatic analysis were written by Astorga, G and were discussed with Tiwari, S.K., and Savva, G., Mather, A.E., Bloomfield, S. J., and de Oliveira Martins, L. The genome wide association study was discussed by Astorga, G., Tiwari, S.K., and Savva, G. Astorga, G., Mather, A.E., Frosini, S.M., Singh, D., Wain, J., Bloomfield, S.J., Savva, G., Tiwari, S.K., and de Oliveira Martins, L. all discussed the work.

2.1.1 Food sample collection

E. coli positive samples were selected from a study described by Janecko and Zamudio *et al.* (Janecko *et al.*, 2023) where a cross-sectional study was conducted in Norfolk, United Kingdom between May 2018 to November 2019 to investigate pathogen contamination by Mather, A.E., Janecko, N., Bloomfield S.J., and Palau, R. The sampling focused on five food commodities that reflected highly consumed food products based on the Family Foods 2015 consumption data (Department for Environment Food and Rural Affairs, 2017). The five food commodities included raw chicken, raw pork, raw/cooked prawns, raw salmon, and pre-packaged leafy greens, resulting in 1,369 food samples (Janecko *et al.*, 2023).

2.1.2 Microbiological detection, processing, and isolation

Sample processing, preparation and *E. coli* microbiological detection were previously described in Janecko and Zamudio *et al.* by Mather, A.E., Janecko, N., Bloomfield S.J., and Palau, R. (Janecko *et al.*, 2023). Samples were processed within 24h from purchase and approximately 100 g of each sample was transferred into stomacher bags, whilst maintaining a sterile chain. All samples were homogenised in 225 mL of buffered peptone water (BPW) (Southern Group Laboratory (SGL), Corby, UK) at 100 rpm for 30 s (Seward stomacher 400C laboratory blender, Worthing, UK), then incubated at 37 °C for 24h ± 3h.

50mL of the incubated BPW was added to 50mL of *E. coli* enrichment broth double concentration (EC 2x) (ThermoFisher Diagnostics, Rochford, UK), and incubated at 42 °C for 24h ± 3h. Eosin methylene blue agar (EMB) (Sigma Aldrich, Haverhill, UK) was inoculated with 10µL loops of the incubated broth and incubated at 37 °C for 24h ± 3h; all subsequent agar plates were incubated for the same time and temperature.

Up to four colonies were selected for further subculturing if they exhibited typical *E. coli* morphology (dark maroon colony). These colonies are hereafter termed isolates. Each isolate was subcultured on MacConkey agar (ThermoFisher Diagnostics, Rochford, UK) and finally subcultured onto tryptic soy agar (Trafalgar Scientific Ltd., Leicester, UK). Isolates were biochemically confirmed to be *E. coli* by a negative result using Simmon's citrate agar (Sigma Aldrich, Haverhill, UK) and a positive result using Remel™ indole spot reagent test (Fisher Scientific, Loughborough, UK). A food sample was classified as *E. coli* positive if there was at least one isolate passed all biochemical tests. Each *E. coli*-positive sample had up to four *E. coli* colonies stored at −70 °C in 1 mL of Brucella broth +17.5% glycerol (ThermoFisher Diagnostics, Rochford, UK).

2.1.3 Food datasets included in the final collection of *E. coli*

In the study from Janecko and Zamudio *et al.*, 859 out of the 1,369 samples were *E. coli* positive (Janecko *et al.*, 2023). Prior to the start of this project, *E. coli* isolates were selected for WGS according to the criteria outlined below.

An initial collection of 516 *E. coli* isolates from 197 *E. coli*-positive samples were collected and sequenced before the start of this project by Mather, A.E., Janecko, N., Bloomfield S.J., and Palau, R. These 516 *E. coli* genomes were used for chapter 3 of this thesis. The samples were collected for four reasons:

1) Chicken (n = 15/15 *E. coli*-positive samples), leafy greens (n = 7/15 *E. coli*-positive samples), pork (n = 10/15 *E. coli*-positive samples), prawns (n = 6/9 *E. coli*-positive samples) and salmon (n = 2/9 *E. coli*-positive samples) samples were previously selected to confirm that the culture method was accurate for detecting *E. coli.* Up to a maximum of four isolates with different morphologies were taken from each *E. coli* positive sample.

2) Chicken (n = 28/28 *E. coli*-positive samples), leafy greens (n = 18/24 *E. coli*-positive samples), pork (n = 14/15 *E. coli*-positive samples), prawns (n = 3/18 *E. coli*-positive samples), and salmon (n = 5/24 *E. coli*-positive samples) samples were previously used to assess host DNA depletion methods for food metagenomes (Bloomfield *et al.*, 2023b) and deposited in the Sequence Read Archive under project PRJNA1107692. Up to a maximum of four isolates with different morphologies were taken from each *E. coli* positive sample.

3) Prawn samples (n = 57/175 *E. coli*-positive samples) were previously chosen to investigate *E. coli* contamination on prawns, where up to a maximum of two isolates were chosen for each *E. coli* positive sample.

4) Chicken (n = 10 *E. coli*-positive samples), leafy greens (n = 8 *E. coli*-positive samples), pork (n = 4 *E. coli*-positive samples) and salmon (n = 10 *E. coli*-positive samples) samples were collected, where up to a maximum of two isolates were randomly chosen for each *E. coli* positive sample.

To explore the diversity of *E. coli* contaminating retail foods in chapter 4, Astorga, G. prioritised samples with four isolates for further sequencing. Bloomfield S.J., and Palau, R. undertook the culturing and growth of the *E. coli*, and the DNA extraction. Further detail for this dataset is described in chapter 4.

2.2 Human and companion animal dataset

To explore *E. coli* co-occurrence between humans and their companion dogs co-habiting within the same household, faecal samples from human and animal hosts were obtained in collaboration with the Royal Veterinary College. Faecal sample collection and *E. coli* detection are described in **section 5.3.1** and **section 5.3.2.** Further details for downstream bioinformatic analysis are stated in chapter 5.

2.2.1 Human and companion dog sample collection

Faecal samples were collected as part of an international longitudinal observational study at the Royal Veterinary College (RVC), Hertfordshire, UK. The design of original survey and sample collection was done by Menezes, J., Pomba, C., Frosini, S.M., and Loeffler, A., in collaboration between the University of Lisbon and the Royal Veterinary College. This study included UK households with at least one human and at least one companion dog living in close contact using convenience sampling. All human participants consented to the main goals of the study. An epidemiological questionnaire was completed by human members of each household (Menezes *et al.*, 2023). Companion animals' health was evaluated by veterinary professionals, which was also supplemented by behavioural data provided by the owner.

An epidemiological questionnaire was completed by at least one of the human members of each household (Menezes *et al.*, 2023), obtaining information regarding the home owner's age, general health, current and past medical issues, travel patterns, and the level of interaction with their companion dogs. Further information is stated in **section 5.3.1.**

2.2.2 Sample collection and processing

The study design for work within this thesis, growth, detection and isolation of the *E. coli* were done by Astorga, G., and Frosini, S.M. Faecal sample collection and processing is described in detail **section 5.3.2**.

2.3 Whole genome sequencing

All presumptive *E. coli* isolates were stored in 17.5% glycerol stocks and frozen at -70ºC were cultured onto tryptic soy agar (Trafalgar Scientific Ltd., Leicester, UK). *E. coli* isolates underwent DNA extraction and sequencing. Growth, culturing, and DNA extraction was performed by Astorga, G. From a pure culture, 2-3 *E. coli* colonies were inoculated into 400µL phosphate buffered saline (PBS) and loaded directly into the Maxwell cartridge. Individual *E. coli* isolate DNA was extracted using Promega Maxwell RSC Cultured Cell DNA kits (Promega, Southampton, UK) according to manufacturer instructions.

Genomes underwent short read sequencing. Nextera Flex DNA libraries were formed using the Nextera XT library preparation kits (Illumina, San Diego, California, USA) and sequenced on an Illumina NextSeq system to create 150bp Illumina paired end reads by the Quadram sequencing team. Whole genome sequenced isolates are hereafter referred to as genomes (Figure 2.1).

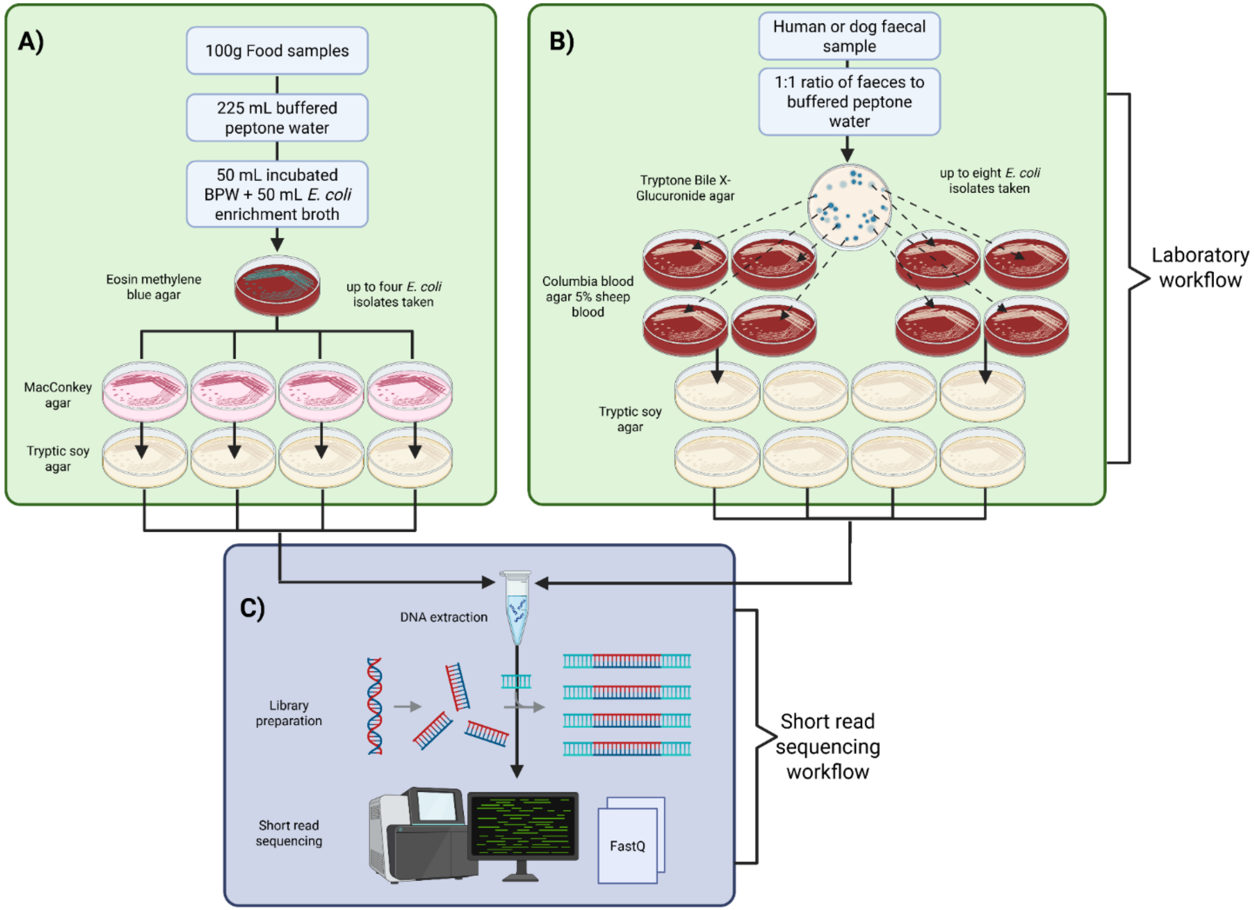


Figure 2.1: Overview of the laboratory workflows for the A) *E. coli* detection and isolation on retail foods, and B) *E. coli* detection and isolation from human and companion dog faeces. *E. coli* collected from both workflows underwent a C) short read sequencing work (Created with BioRender.com).

2.4 Genome analysis

Genomic analyses were performed on the High-Performance Computing (HPC) cluster at the Norwich Biosciences Institutes on the Quadram Institute Bioscience QIB-CLOUD Virtual Machine and Galaxy web platform (Afgan *et al.*, 2016). Genomic analyses that were consistent across all chapters are outlined here in Chapter 2. Separate analyses unique to each chapter are defined in each respective chapter. The data analysis, scripts and code were written by Astorga, G. with guidance from Zamudio, R., and de Oliveira Martins, L., Bloomfield, S.J., Singh, D., Zamudio, R., Tiwari, S.K., Savva, G., and Mather, A.E.

2.4.1 Quality control and assembly

Paired raw read adapter sequences were trimmed using Trimmomatic v.0.33 using the LEADING:3 TRAILING:3 SLIDINGWINDOW:4:25 MINLEN:36 parameters (Bolger *et al.*, 2014). Genomes were classified into presumptive species using Centrifuge v.1.0.3 (Kim *et al.*, 2016), where trimmed raw reads were compared against a Centrifuge bacterial genome reference database downloaded on July 14, 2022 .

Trimmed reads were assembled using SPAdes v.3.1.1 (Bankevich *et al.*, 2012). Assembly quality was analysed using QUAST v.4.6.3 (Gurevich *et al.*, 2013) and by aligning the trimmed reads to the genome assemblies using the Burrows-Wheeler aligner v.0.7.17 (Li and Durbin, 2009), and analysing the read depth and coverage using samtools v1.9 (Li *et al.*, 2009) and bcftools v.18 by SAMtools (Li *et al.*, 2009). Assembly quality was checked by CheckM v.1.0.11 (Parks *et al.*, 2015), whereby assemblies with >50 duplicate marker genes were re-isolated and re-sequenced.

The average nucleotide identities of the genome assemblies were compared to *E. coli* K-12 substr. MG1655 (Blattner *et al.*, 1997) (Genbank Accession assembly ID: GCA_000005845.2) using FastANI v.1.32 (Jain *et al.*, 2018). *E. coli* genomes passed quality control if they were 4.4 Mbp – 5.7. Mbp in length, had an average GC content between 42.6%-54.6% (Deneke *et al.*, 2021), and had a mean read depth of their four largest contigs above 28.

2.4.2 *E. coli* phylogroup and sequence typing

E. coli genome assemblies were classified into phylogroups using Clermontyping v.20.03 (Beghain *et al.*, 2018), and into sequence types (STs) using MLST v.2.19.0 (Seemann, 2022) with the *E. coli* pubMLST seven gene scheme (Seemann, 2022). Unclassified STs were further characterised into matching novel STs by identifying unique combinations of *adk, fumC, gyrB, icd, mdh, purA,* and *recA* profiles provided by the MLST output.

2.4.3 Antimicrobial resistance determinants and virulence genes

The presence of AMR genes, plasmid replicons and virulence genes were determined using ARIBA v.2.14.6 (Hunt *et al.*, 2017) with default settings on the trimmed reads. ARIBA was used alongside the Resfinder (Bortolaia *et al.*, 2020), PlasmidFinder (Carattoli and Hasman, 2020), and VFDB (virulence factor database) core (Chen *et al.*, 2016) databases that were downloaded in May 2023 (Figure 2.2). Point mutations associated with AMR were identified using PointFinder (Zankari *et al.*, 2017). AMR genes and point mutations were collectively termed AMR determinants.

The AMR determinants present were categorised into their antimicrobial classes and priority according to the World Health Organization scheme (World Health Organization, 2024). Putative multidrug-resistance (MDR) in this study was classified as the presence of AMR genes or point mutations conferring resistance to three or more different antimicrobial classes according to the WHO antimicrobial class list (World Health Organization, 2019, Catalano *et al.*, 2022).

E. coli isolates were characterised as extraintestinal pathogenic (ExPEC) lineages if two or more of the following virulence genes were present in the genome: *papAH, papC, sfa/focDE, afa/draBC, iutA,* and *kpsMII* (Johnson *et al.*, 2003).

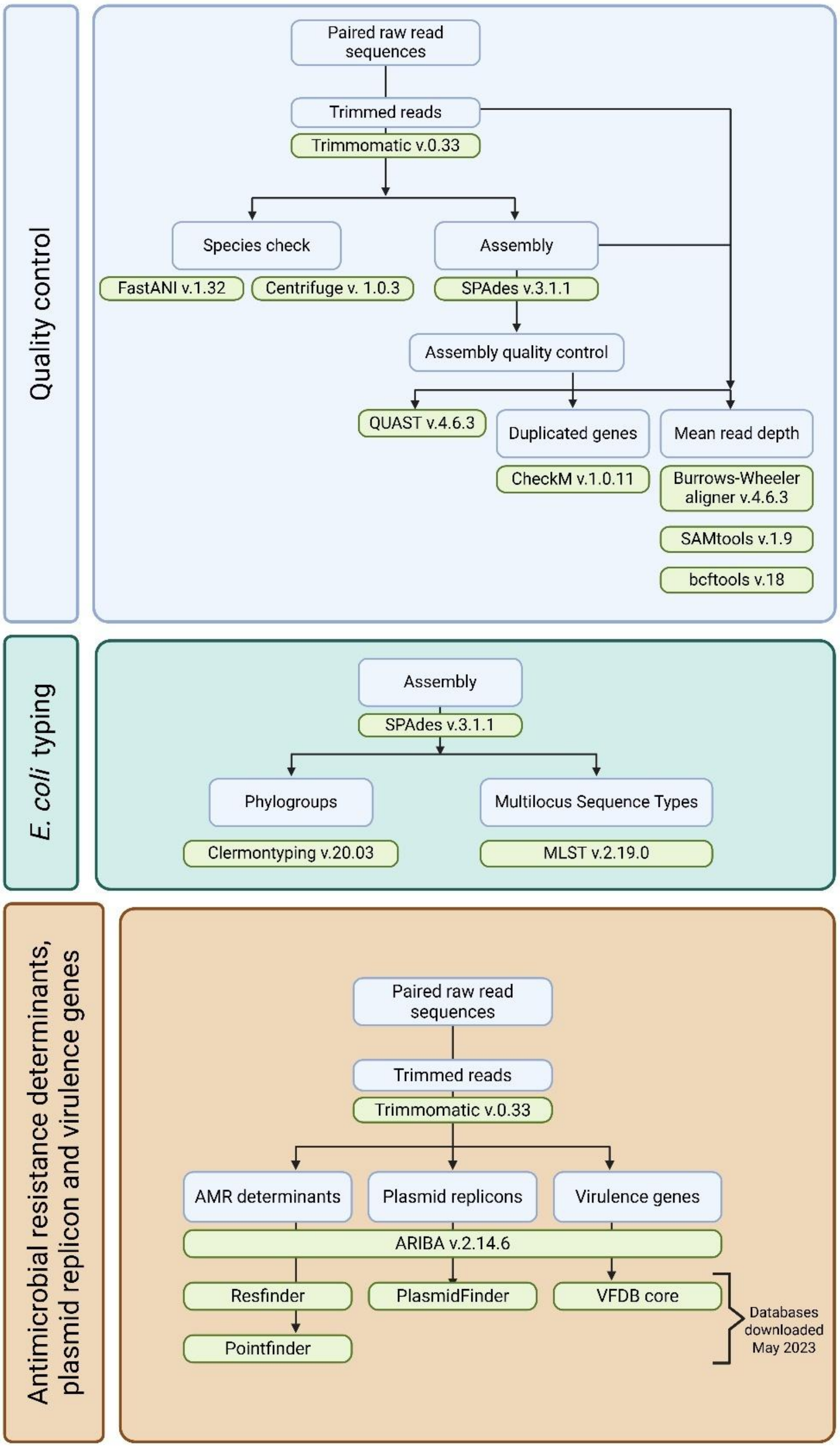


Figure 2.2: Summary workflow of the short-read sequence quality control processes, typing and identification of antimicrobial resistance determinants, plasmid replicons and virulence genes. All versions of the tools and the dates when databases were downloaded are specified.

2.5 Data analysis, statistical analysis, and visualisation

All statistical analyses and visualisation were performed in R v.4.1.3 in RStudio v. 2021.09.1, and subsequent updated versions are stated in each chapter. All packages used are stated in each chapter. All statistics analysis was done by Astorga, G. with guidance from Zamudio, R., and de Oliveira Martins, L., Bloomfield, S.J., Singh, D., Zamudio, R., Tiwari, S.K., Savva, G., and Mather, A.E.

Chapter 3: Comparison of phylogenomic approaches on *E. coli* populations isolated from retail foods

Chapter contributions: Design of the food survey and sample collection was done by Mather, A.E., Janecko, N., Bloomfield S.J., and Palau, R. Collection of the food samples, as well as growth, detection and isolation of the *E. coli* were done by Bloomfield S.J., and Palau, R. Comparison of phylogenomic methods of *E. coli* from retail foods in Norfolk, United Kingdom was planned and developed by Astorga, G. and de Oliveira Martins, L; Bioinformatics was taught and facilitated by de Oliveira Martins, L. and Zamudio, R; Scripts and code were written by Astorga, G. and de Oliveira Martins, L; Astorga, G., Mather, A.E., Singh, D., Wain, J., Zamudio, R., and de Oliveira Martins, L. all discussed the work.

Publication: This chapter is in preparation for submission in Access Microbiology under the working title "Comparison of phylogenomic approaches on *Escherichia coli* isolated from retail foods"

3.1 Introduction

As described in chapter 1, phylogenetics is an umbrella term for the branch of research of the evolutionary history and genetic relationships among a group of organisms. The main goal of phylogenetics is to infer the most accurate evolutionary history (Wang *et al.*, 2015). This can be done by comparing the physiological features (Semple and Steel, 2003) or comparison of a select few genes in the group of organisms, for instance. Phylogenomics is an area of phylogenetics which compares the whole genome or large parts of the genomes when inferring the evolutionary history. Phylogenomics is the sub-branch of phylogenetics that expands traditional phylogenetics by taking into account differing evolutionary mechanisms and large amounts of genomic information (Rodríguez-Ezpeleta and Philippe, 2009).

E. coli was initially thought to have a clonal structure (Tenaillon *et al.*, 2010), but whole genome sequencing has shown *E. coli* to be a genetically diverse organism that can rearrange genetic material through recombination, and acquire new genetic material from other microorganisms through horizontal gene transfer (Guttman and Dykhuizen, 1994). The rate of recombination has been found to be higher than the rate of mutation in certain regions of the *E. coli* genome, which can make it difficult for inferring the phylogenetic tree (Arenas *et al.*, 2008). Additionally, horizontal gene transfer is common in *E. coli* (Lawrence and Ochman, 1998), and has been known to interfere with inferring the evolutionary relationship between isolates (Dykhuizen and Green, 1991).

Phylogenomic approaches are required to obtain a more accurate understanding of the evolutionary relationships between different isolates of *E. coli*. Phylogenomic approaches have surpassed their phylogenetic predecessors, with previous work stating that they have become the gold standard (Steenwyk *et al.*, 2023). However, there are many phylogenomic approaches and there is currently no standard on which approach to take. Each approach utilises the genomes in different ways and makes different assumptions about the dataset.

3.2 Phylogenomic methods

In this chapter, five approaches were investigated. These were categorised into sequence-based phylogenomic approaches and whole genome feature-based approaches (Figure 3.1). It is currently not clear how these approaches fare against the same dataset and whether there is one that is the most suitable for inferring which *E. coli* are from a specific source.

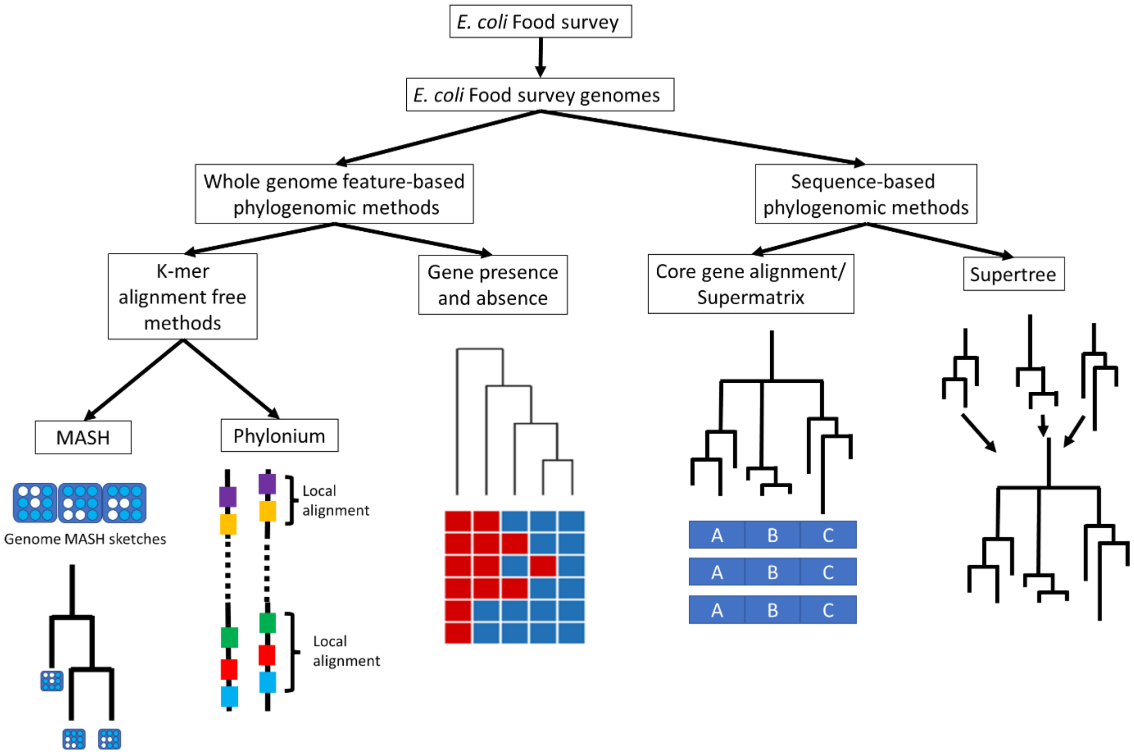


Figure 3.1: Overview of the five phylogenomic methods used in this study, adapted from Delsuc et al. (2005) (Delsuc *et al.*, 2005). Whole genome feature-based phylogenomic methods are split into the gene presence and absence approach, and the k-mer alignment free approaches: MASH (Ondov *et al.*, 2016) and Phylonium (Klotzl and Haubold, 2020). Sequence-based phylogenomic methods are split into the Core gene alignment/Supermatrix and Supertree approaches.

3.2.1 Whole genome feature-based approaches

In this chapter, whole genome feature-based approaches are defined as approaches that utilise the whole genome via the presence and absence of genes, or nucleotide sequences at length ‘k’ (k-mers) to infer the genetic similarity between genomes.

The gene presence/absence approach uses the concept that more closely related genomes will share more genes. Orthologs are genes that come from an ancestor which have then evolved through speciation events. All the genes in a group of *E.*

coli are not necessarily conserved, and this collection of conserved and non-conserved genes is termed the pangenome. The gene presence/absence approach uses the presence and absence of pangenome orthologous groups (POGs) to infer the genetic similarity between genomes. A binary presence/absence matrix is made using these POGs, which reflects how similar the genomes are (Page *et al.*, 2015).

Alignment-free approaches use k-mers, which are short nucleotide sequences at a fixed length ‘k’. These approaches are argued to be more robust to genome rearrangement and lateral gene flow (Zielezinski *et al.*, 2017). There are several alignment free methods for genome comparison, which include word counts (Sims and Kim, 2011), average longest common k-mers (Ulitsky *et al.*, 2006), chaos theory-based sequence representation (Almeida *et al.*, 2001), and iterated maps (Almeida, 2013). These tools use variation between exact matching sequences and are converted into distance matrices without using sequence alignments.

Both the gene presence/absence method and the alignment free approaches calculate the genetic similarity between genomes and create a distance matrix. There are different methods to construct a dendrogram from these matrices, including optimality-based methods and clustering algorithms. The optimality-based methods compare different tree topologies and select the topology that seems to be the best fit, according to the least-squares or the minimum evolution criteria. There are also commonly used clustering algorithms, like the neighbour-joining method or the unweighted pair group method with arithmetic mean (UPGMA) clustering method which assume a constant evolutionary rate and that all sequences were sampled at the same time (Yang, 2014).

3.2.2 Sequence-based phylogenomic approaches

Sequence-based phylogenomic approaches, as defined here, utilise, and compare aligned homologous sections of the genome at the nucleotide level from whole genome sequences. A common phylogenomic approach for inferring the genomic relationship between *E. coli* genomes is the core gene alignment approach. The core gene alignment approach compares homologous sites between concatenated sequence alignments. However, there are many ways to create these multiple sequence alignments and currently no standard for the overall core gene alignment approach.

E. coli is an example of an organism with an open pangenome, where new genes can generally be found when a new *E. coli* isolate is added to a dataset (Touchon *et al.*, 2009). Genes that are defined as "core", are present in all genomes within a given set of genomes. The term "core gene" is not always the same in every study, as the term is sometimes used to refer to genes present in a select percentage of the population (Xia *et al.*, 2022, Kaas *et al.*, 2012). Additionally, the "core genes" defined in these studies refer to a computational definition of core genes, where these genes are selected from their respective datasets. These "computational core genes" may not be the exact same as the biological core genes that are essential for all *E. coli* function and survival (Gerdes *et al.*, 2003). In this chapter, when "core genes" are referred to, they reference the computationally selected core genes.

One example of a core gene alignment approach is the "core gene alignment" method. The "core genes" from each individual genome are selected and aligned to the other homologous genes from other genomes. This creates a gene alignment, where each row is the same gene from a different genome, and each column is a homologous site. These individual gene alignments can be merged to the other gene alignments, creating a matrix where each row is now an alignment of the core genes from individual genomes. The homologous sites in this final alignment are compared using a standard phylogenetic inference procedure like maximum likelihood (Shakya *et al.*, 2020).

The other sequence-based phylogenomic method is termed the Supertree approach (Daubin *et al.*, 2001). A Supertree is constructed by first creating unrooted gene trees. These gene trees are used to create a species tree. One way this can be done is by dividing the gene trees into unrooted quartets, which are four-taxon trees. These quartets are used to construct a Supertree with the least amount of disagreement. Supertree methods have been used to reconstruct bacterial phylogenies (Daubin *et al.*, 2001, Lang *et al.*, 2013), eukaryotic species and viruses (Lang *et al.*, 2013). It has been found with simulated data that

Supertree approaches are more reliable when dealing with moderate amounts of horizontal gene transfer (Lapierre *et al.*, 2014).

3.3 Aims and Objectives

These phylogenomic methods differ by the number of steps to produce the final tree, the amount of time taken to produce this tree and the inference of the evolutionary relationship between isolates. As there is no standard phylogenomic approach, five different phylogenomic methods were evaluated against the same *E. coli* genomes isolated from different food products, comparing the inferred population structures and tree topologies produced. The main goal of this study was to identify which phylogenomic approach would produce an efficient result that would be the best suited for describing the population. The selected approach could then be applied to further analyses within this thesis.

The hypothesis for this chapter was that different phylogenomic methods ((Roary - core gene alignment), gene presence and absence, alignment-free (MASH), alignment-free (Phylonium) and Supertree (ASTRAL-pro)) will generate different tree topologies and provide different insights into the relationships of the *E. coli* within this dataset's population.

The hypothesis will be explored by addressing the following questions:

1) will the core gene alignment, gene presence and absence approach, alignment free approaches and the Supertree approach show different population structures?
2) Do the different approaches create different topologies based on the same population dataset?
3) Determine which approach would be the best suited for describing what types of *E. coli* are present from different sources.

3.4 Materials and methods

3.4.1 Dataset of *E. coli* whole genome sequences

The collection, quality control and phylogroup classification of the *E. coli* used in this chapter is described in chapter 2 of this thesis. The investigation into the diversity of *E. coli* on retail foods is described in Chapter 4. Briefly, the dataset for this chapter consists of *E. coli* collected from a food survey in Norfolk, United Kingdom described in Janecko, Zamudio *et al*. (Janecko *et al.*, 2023). *E. coli* were isolated from five food sources (chicken, leafy greens, pork, prawn, and salmon). *E. coli* was cultured from chilled, frozen whole chicken or chicken pieces; fresh pre-packaged leafy greens and packaged whole salad greens; chilled or frozen pork muscle cuts, chilled or frozen (whole/shelled) prawns; and chilled or frozen salmon fillets or steaks. Up to four colonies (hereafter referred to as isolates) were selected from an individual food sample. The total of 516 *E. coli* isolates included those from chicken (n = 175), leafy greens (n = 88), pork (n = 84), prawns (n = 129) and salmon (n = 40) (Table 3.1). Each isolate from the same sample was treated as an independent organism.

Table 3.1: Total number of *E. coli* genomes in each sample in the food survey dataset, separated by food commodity.

Commodity	1 genome in a sample	2 genomes in a sample	3 genomes in a sample	4 genomes in a sample	Total number of genomes
Chicken	9	1	8	35	175
Leafy greens	11	2	7	13	88
Pork	6	2	6	14	83

Table 3.1: Total number of *E. coli* genomes in each sample in the food survey dataset, separated by food commodity.

Commodity	**1 genome in a sample**	**2 genomes in a sample**	**3 genomes in a sample**	**4 genomes in a sample**	**Total number of genomes**
Prawn	17	41	2	6	129
Salmon	5	5	3	4	40

As cryptic clades of *E. coli* are known to produced elongated branch lengths in phylogenies, any classified cryptic clades were removed from phylogenetic analyses. One cryptic clade was present in this dataset was removed from the phylogenetic analyses. The remaining 515 *E. coli* isolates were analysed using five different phylogenetic approaches (Figure 3.2).

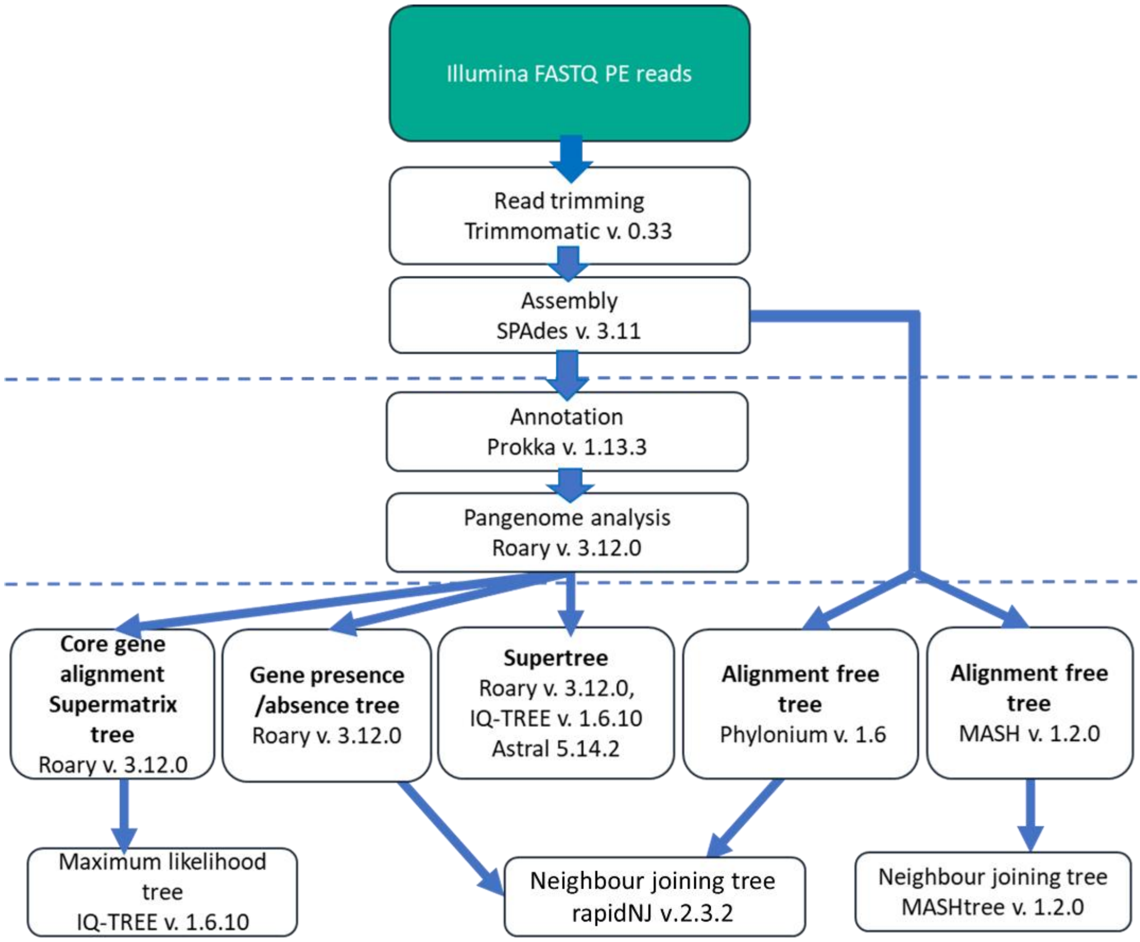


Figure 3.2: Bioinformatic workflows of the five different phylogenomic approaches used in this chapter.

3.4.2 Gene presence and absence

A gene presence and absence tree using Roary v.3.12.0 (Page *et al.*, 2015) was created with the -a -cd 95 -I 90 -p 32 -g 100000 -e -n parameters on the 515 *E. coli* isolates. Roary creates a binary matrix of present and absent genes. The Jaccard coefficient was used to calculate the values for the similarity matrix of present and absent genes between genomes. The resulting similarity matrix was then clustered by the UPGMA algorithm to create a dendrogram. A dendrogram was inferred using rapidNJ v.2.3.2 (Simonsen *et al.*, 2008) with the -i pd -o t -x options (Figure 3.3).

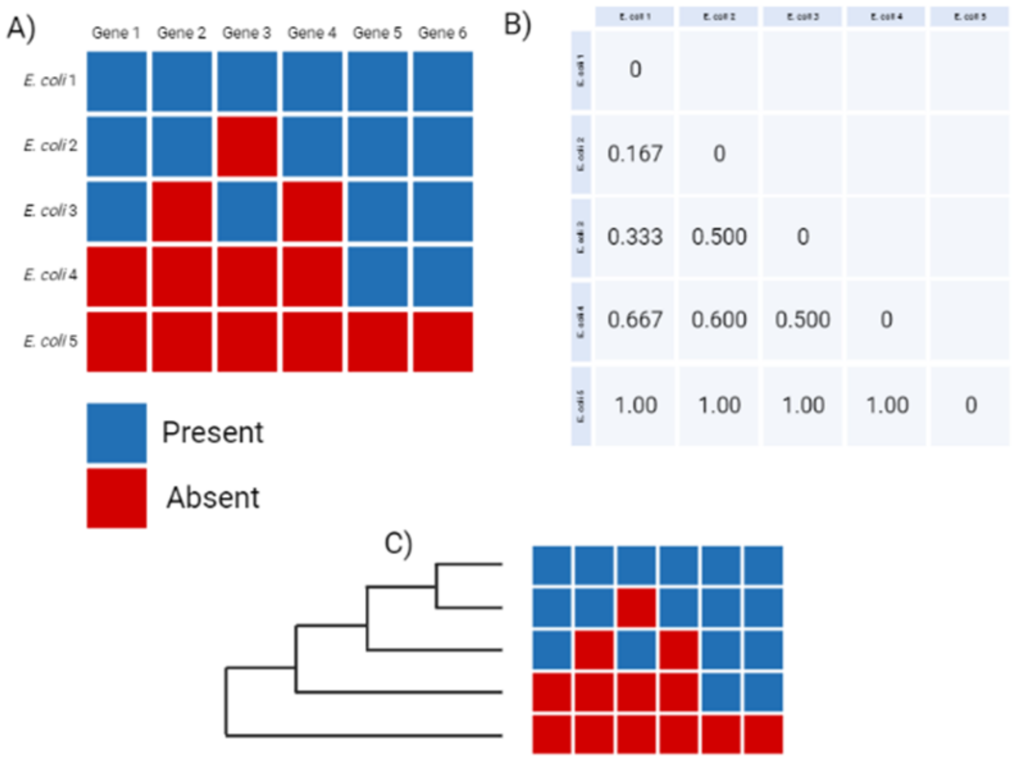


Figure 3.3: A simplified workflow showing how a dendrogram can be created from the presence and absence of genes. (A) The presence and absence matrix of genes, (B) the distance matrix, and (C) showing the dendrogram.

3.4.3 Alignment-free approaches

Zielezinski *et al*. (2019) compared 74 alignment-free approaches and concluded that MASH was one of the best performing software on simulated data, whilst Phylonium was tied first with Andi and co-phylog on real-world sequence data (Zielezinski *et al.*, 2017). Phylonium was found to be the best performing software on *E. coli* datasets. Therefore, MASH and Phylonium were chosen as the approaches to take forward for comparison in this analysis.

MASH v.2.0 (Ondov *et al.*, 2016) estimates the evolutionary relationships using the MinHash algorithm to turn k-mers into smaller identifiable “hashes”. Hashes are sorted and a subset of the top hashes are turned into a “sketch”. These sketches are sorted by size and the smallest hash sketches (min-hashes) are compared between isolates using the Jaccard index to find the fraction of shared hashes (Katz *et al.*, 2019). MASH v2.0 then created a distance matrix (Figure 3.4). Genetic distances between the 515 *E. coli* genomes were calculated using default parameters. Genetic distances are recorded as a pairwise distance matrix and used to create a tree using a neighbour joining algorithm via MASHtree v.1.2.0 (Katz *et al.*, 2019) with the –mindepth 0 parameter.

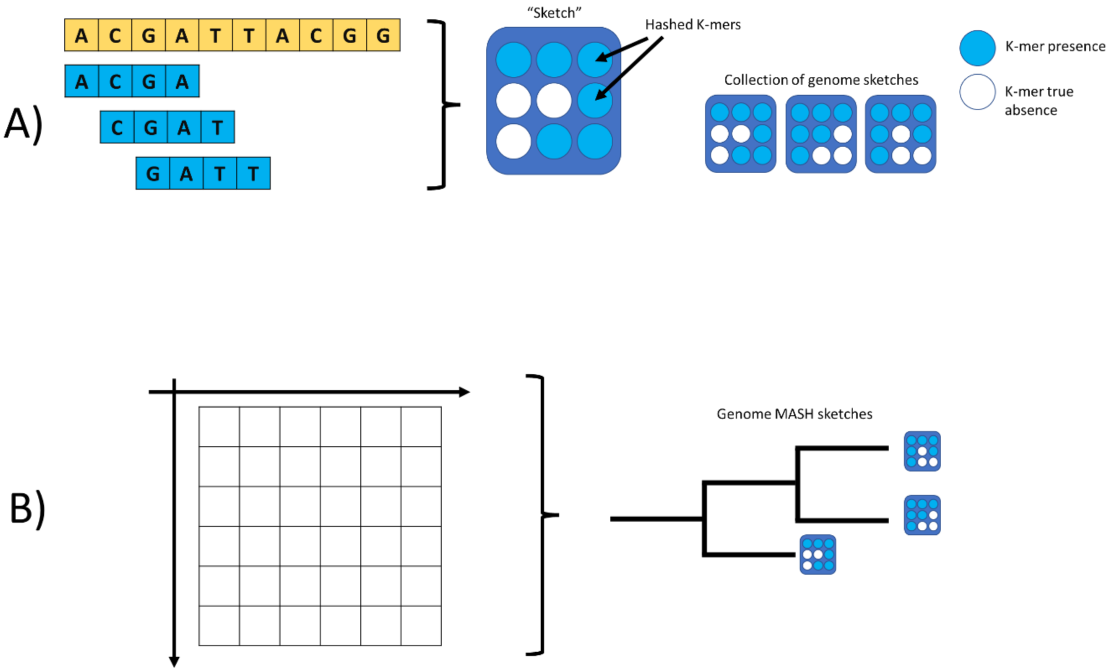


Figure 3.4: A simplified workflow to estimate MASH (Ondov *et al.*, 2016) distances between genomes and create a phylogenomic dendrogram. A) Genomes are split into k-mers at a set length specified by the parameters, transformed into hashes, sorted and then subset into a sketch. B) The k-mer presence and absence is used as input for a distance matrix, which is then turned into a dendrogram using the a neighbour-joining algorithm via MASHtree (Katz *et al.*, 2019).

Phylonium v.1.6 (Klotzl and Haubold, 2020) differs from MASH as Phylonium selects a reference genome from the dataset and find matching “anchor” k-mers and uses these equidistant anchors to find and distinguish homologous regions. The SNPs are counted in these homologous regions and used to estimate the evolutionary relationship (Figure 3.5). Phylonium v.1.6 was used to estimate the distances between the 515 *E. coli* genomes using default parameters. Phylonium created a distance matrix in PHYLIP format and a dendrogram was estimated using rapidNJ v.2.3.2 (Simonsen *et al.*, 2008) with the -i pd -o t -x options.

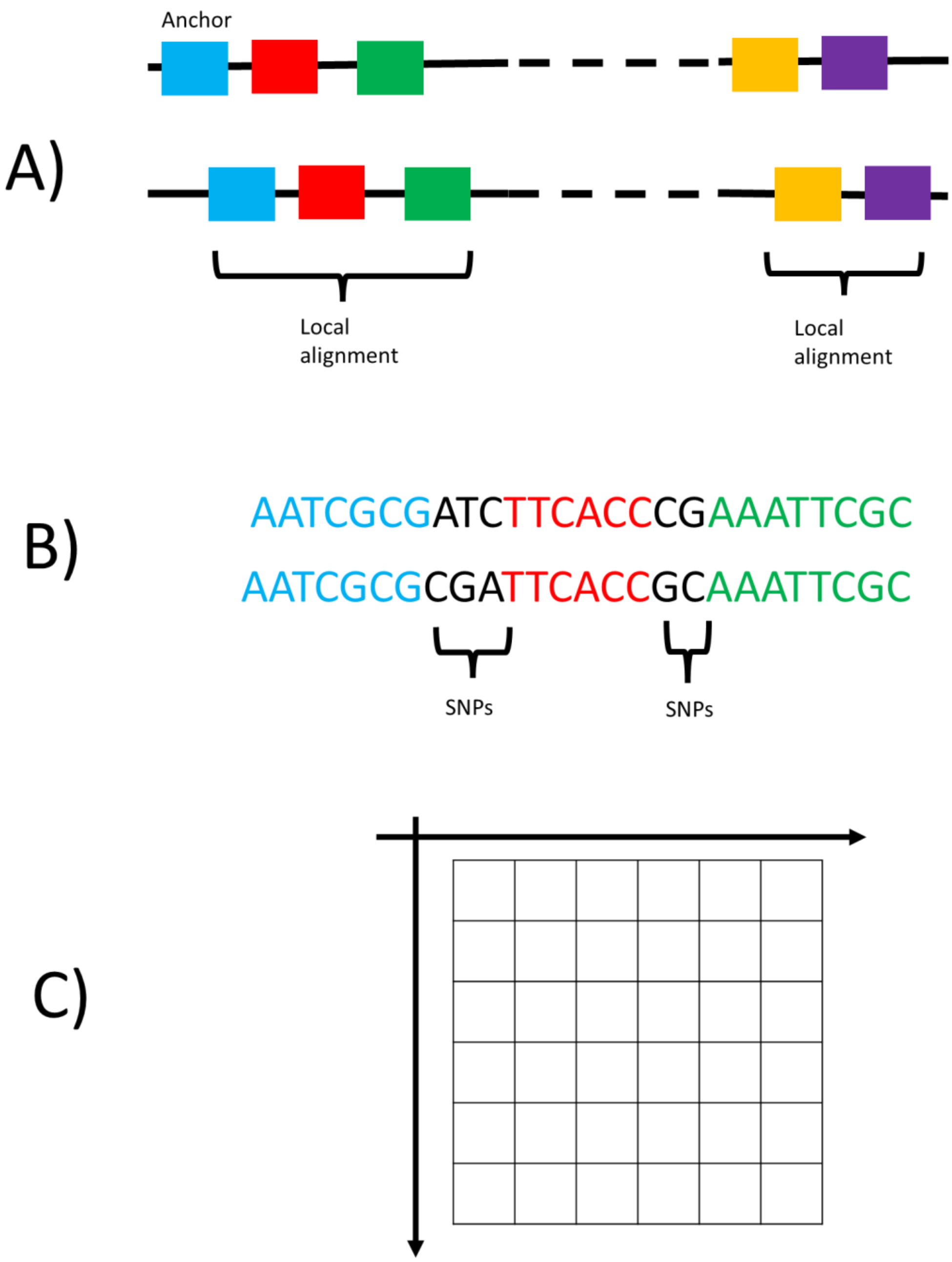


Figure 3.5: A simplified explanation of Phylonium. A) Whole genome sequences are used as inputs, where homologous regions are determined by finding equidistant k-mers termed “anchors”. B) SNPs are found within these regions. C) A distance matrix is calculated from the SNP regions and is used to create a phylogenomic tree (Created with BioRender.com).

3.4.4 Core gene alignment

Pangenome analysis was performed using the annotated draft genomes and Roary v.3.12.0 (Page *et al.*, 2015) using the -a -cd 95 -I 90 -p 32 -g 100000 -e -n parameters on the 515 *E. coli* isolates, without the cryptic clade. Core genes were defined as 90% sequence identity similarity for genes in 95% of the isolates. These genes were then aligned to create a core gene alignment (Figure 3.6). The whole core gene alignment was used to construct a maximum-likelihood phylogenetic tree under a Hasegawa, Kishino, and Yano (HKY) substitution model with the Gamma rate heterogeneity model using IQ-TREE v.1.6.10 with the -m HKY+G -nt AUTO parameters.

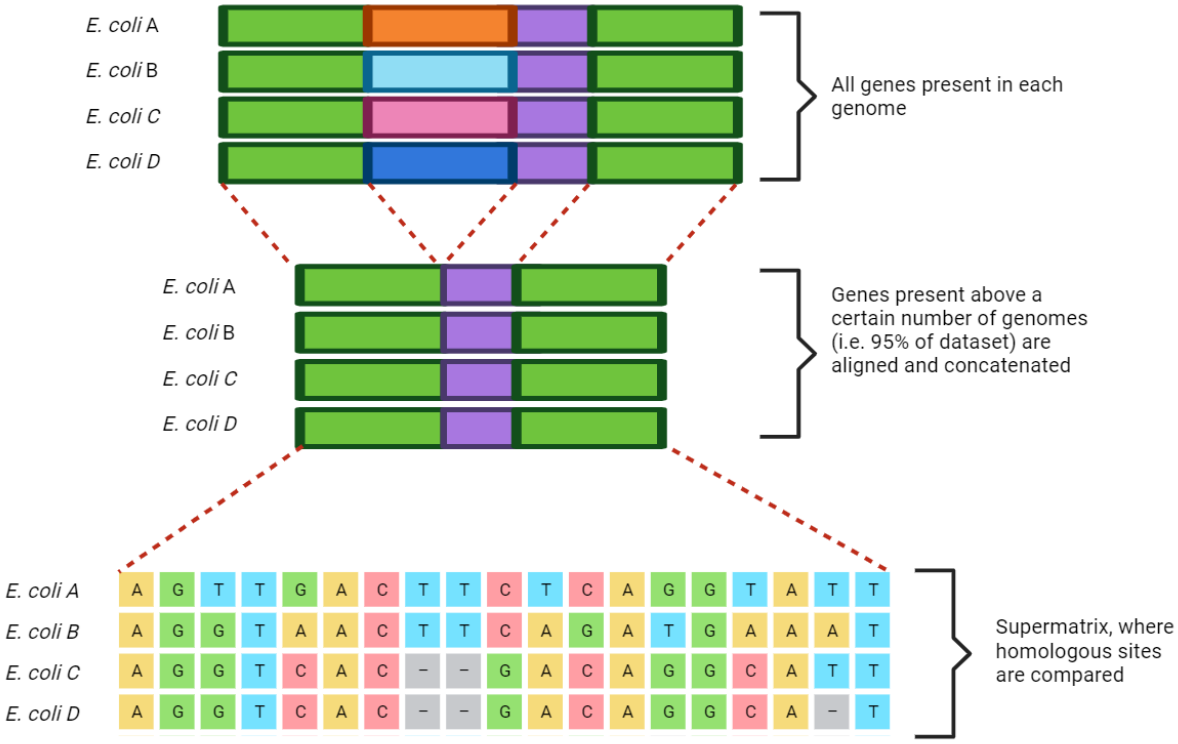


Figure 3.6: A simplified description of Roary (Page *et al.*, 2015) where genes present in 95% of bacterial isolates are selected, aligned and concatenated to create a core gene alignment. Homologous sites are subsequently compared to infer a phylogeny (Created with BioRender.com).

3.4.5 Supertree

Roary v.3.12.0 (Page *et al.*, 2015) created intermediate alignment files of all the annotated gene sequences using the -z parameter. This includes both homologous genes (genes that arise from speciation) and paralogous genes (genes that arise from gene duplication). Each separate alignment file was used to create an individual maximum-likelihood gene tree by using IQ-TREE v.1.6.10 (Nguyen *et al.*, 2015) with the -m HKY+G -nt AUTO parameters. Gene trees were selected for use if they had at least four tips on the tree.

Gene tree statistics (i.e. edge length, nodes and tips) was evaluated using the "stats" option in gotree v.0.4.3 (Lemoine and Gascuel, 2021). ASTER v.1.3 was used to implement ASTRAL-Pro v. 1.4.1.1 (Zhang *et al.*, 2020). ASTRAL-Pro was used to infer a species tree from gene trees created from both orthologs and paralogs.

3.4.6 Tree visualisation

All trees were visualised using R v.4.1.3 (R Core Team, 2022) in RStudio v. 2021.09.1 (RStudio Team, 2020) using the ggtree v.3.2.1 (Yu, 2020), ape v.5.5 (Paradis and Schliep, 2019), phytools v.1.0-1 (Revell, 2012), and ggnewscale v.0.4.5 (Campitelli, 2022) packages.

3.4.7 Determining the standard approach for comparison

The most commonly used phylogenomic approach to evaluate *E. coli* populations was investigated on PubMed. The term "phylogenetic" was used as the umbrella term for the branch of research of the evolutionary history and genetic relationships among a group of organisms. On PubMed (date accessed 07/08/2023), the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population) were used as input and was further filtered for studies published from 2020 onwards. The most commonly used approach and software was chosen to be the standard approach for which all other approaches would be compared to (Appendix 1).

3.4.8 Topology comparisons

Bayesian analysis of population structure (BAPS) rhierBAPS v.1.1.3 R package (Tonkin-Hill *et al.*, 2018, Cheng *et al.*, 2013) was performed on the core gene alignment output from Roary using the max.depth = 2, n.pops = 103, quiet = TRUE, assignment.probs = TRUE, n.cores= 16 parameters. The n.pops value was calculated from the number of isolates divided by five (Tonkin-Hill *et al.*, 2018).

The output produces two levels of population clusters according to their algorithm. To evaluate how well the level 2 core gene alignment BAPS clusters mapped onto the different phylogenies, the silhouette scores for the level 2 BAPS clusters were calculated for each tree.

The 3rd upper quartile and median silhouette score were calculated for each tree. The Robinson-Foulds distance (RF distance) evaluates the number of unique branches between two trees, which allows evaluation of the tree topology. The RF distance is found by calculating the number of partitions that are unique between two trees (i.e. Tree 1 and Tree 2) (Figure 3.7A). The partitions refer to when an internal edge is removed, creating separated tips on a tree (Figure 3.7B). This is defined by Steel and Penny (1993) (Steel and Penny, 1993) and Robinson and Foulds (1981) (Robinson and Foulds, 1981), as the number of partitions that are unique to the first tree (Tree 1) and the number of partitions unique to the second tree (Tree 2) (Figure 3.7C), divided by 2.

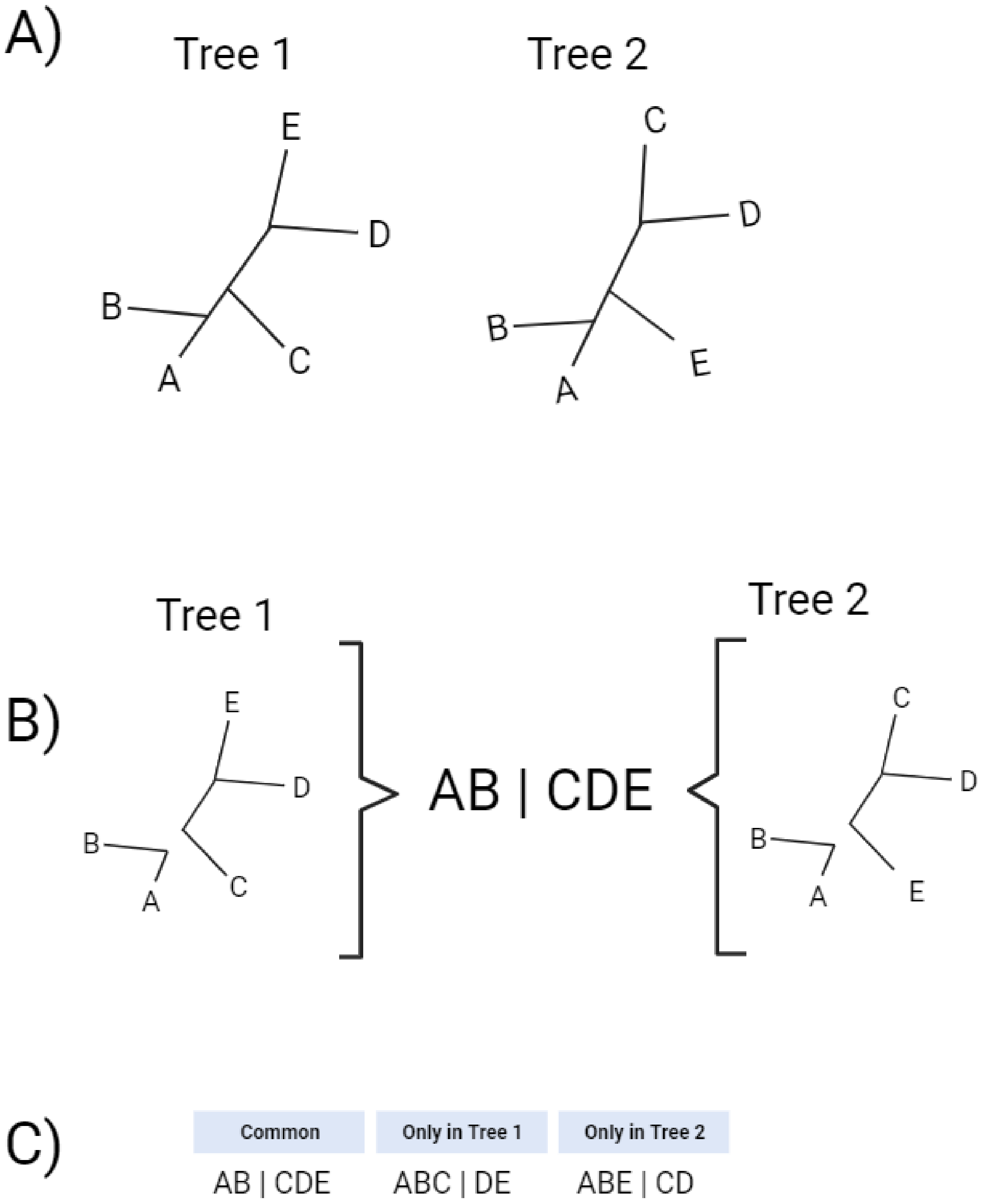


Figure 3.7: (A) Shows an example of two different trees with tips arbitrarily labelled A-E for reference. (B) Indication of the notation used when an internal edge is removed, creating separated tips on a tree (partitions). (C) A summary of the comparison of partitions between the trees, where a single partition is unique to each tree. The sum of the number of partitions unique to the first tree and unique to the second tree, divided by two is the Robinson-Foulds distance.

As the Supertree and the other phylogenomic tree methods use different scales, the branch lengths were made comparable by normalising the values (Equation 1).

The trees were normalised so that the sum of the branch lengths equals one. To do this, the edge lengths ($\boldsymbol{l}$) were divided by the sum of all the edge lengths ($\sum \boldsymbol{l}$) (Didelot and Wilson, 2015).

For example, the core gene alignment edge lengths correspond to nucleotide changes over time, whilst the branch lengths of the Supertree are estimated by ASTRAL-Pro for both internal and terminal branches. Branch lengths for individual gene trees are in substitutions per site and the branch lengths for the Supertree are measured in coalescent units (Zhang *et al.*, 2020, Zhang *et al.*, 2018). Some metrics to determine similarity between tree topologies put weight on long branch lengths more heavily than others. This will cause the topology comparison metrics to evaluate them as extremely different, whereas, in reality, the phylogeny may be similar but in a different scale. Therefore, the branch lengths were normalised such that their proportional lengths could be compared.

$$\text{Normalised branch length} = \frac{l}{\sum l} \qquad \textit{Eq. 1}$$

The weighted Robinson-Foulds distance (wRF) is an extension of the RF distance as defined by Robinson and Foulds (1979) (Robinson and Foulds, 1979). The wRF distance takes the RF distance (the number of unique branches) and the branch lengths into account, where branch length refers to the number of mutations between sequences it connects. The wRF is calculated as the sum of the modulus of the branch lengths of all the matching partitions for the first tree minus the matching partitions for the second tree. In the case that the partition is unique for a tree, the missing branch length is set to 0 (Robinson and Foulds, 1979).

The Kuhner-Felsenstein measure (KF distance), also termed the branch score distance, focuses on the branch lengths of the tree and is determined by calculating the Euclidean distance. This is defined as the sum of the squares of the differences between branch lengths (Kuhner and Felsenstein, 1994).

The topology of the trees is utilised implicitly by considering branch lengths of non-existent splits as 0, as with the wRF distance. In more detail, the KF distance is computed by taking all possible branches in the two trees into account (Kuhner and Felsenstein, 1994). Each partition in the first tree and the second tree has an associated branch length. The difference between the branch lengths is calculated for each matching partition. In the event of one tree not having a particular branch, the difference is between the branch length and 0. The sum of all the differences is calculated and squared, which is then square rooted to provide the branch score (Kuhner and Felsenstein, 1994).

3.5 Results

When the cryptic clade was excluded from analysis, the *E. coli* pangenome for the 515 *E. coli* dataset consisted of 40,921 genes, 7.3% (n = 2,991) of which were genes present in at least 95% of genomes (Figure 3.8). When this core gene alignment was compared to the reference genome *E. coli* K-12 substr. MG1655, the core gene alignment length (2,748,251 bp) represented 59.2% of the reference genome length (4,641,652 bp) (32) (Genbank Accession assembly ID: GCA_000005845.2).

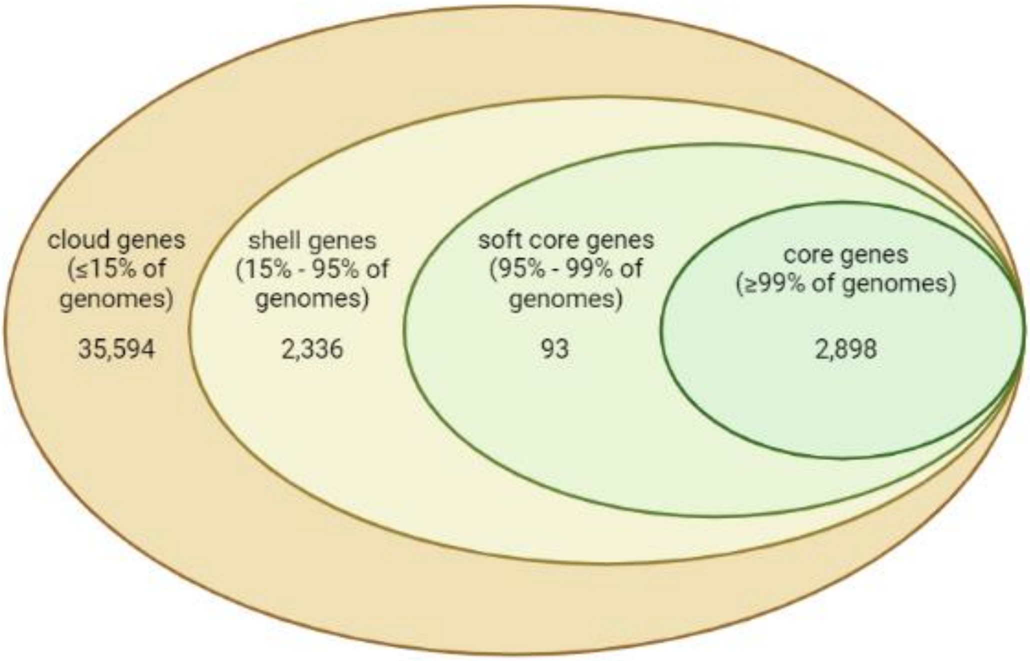


Figure 3.8: The number of core genes (present in ≥99% of the dataset), soft core genes (95%-99% of the dataset), shell genes (15% - 95% of the dataset), and cloud genes (≤15% of the dataset) in the 515 *E. coli* genomes analysed in this study using Roary v.3.12.0 (Page *et al.*, 2015).

3.5.1 Supertree statistics

IQ-TREE took 37 days to complete all trees. Gene trees were used to infer the species tree using ASTRAL-Pro. The sum of the branch lengths for 22,008 single-copy and multi-copy gene trees derived from 515 *E. coli* genomes with at least four tips were used to create the Supertree. The gene trees with at least four tips were found to have a range of tips between 4 – 2613 (median = 14), and the range of the log transformed sum branch lengths were between -5.82 – 1.82 (median = -1.17) (Figure 3.9).

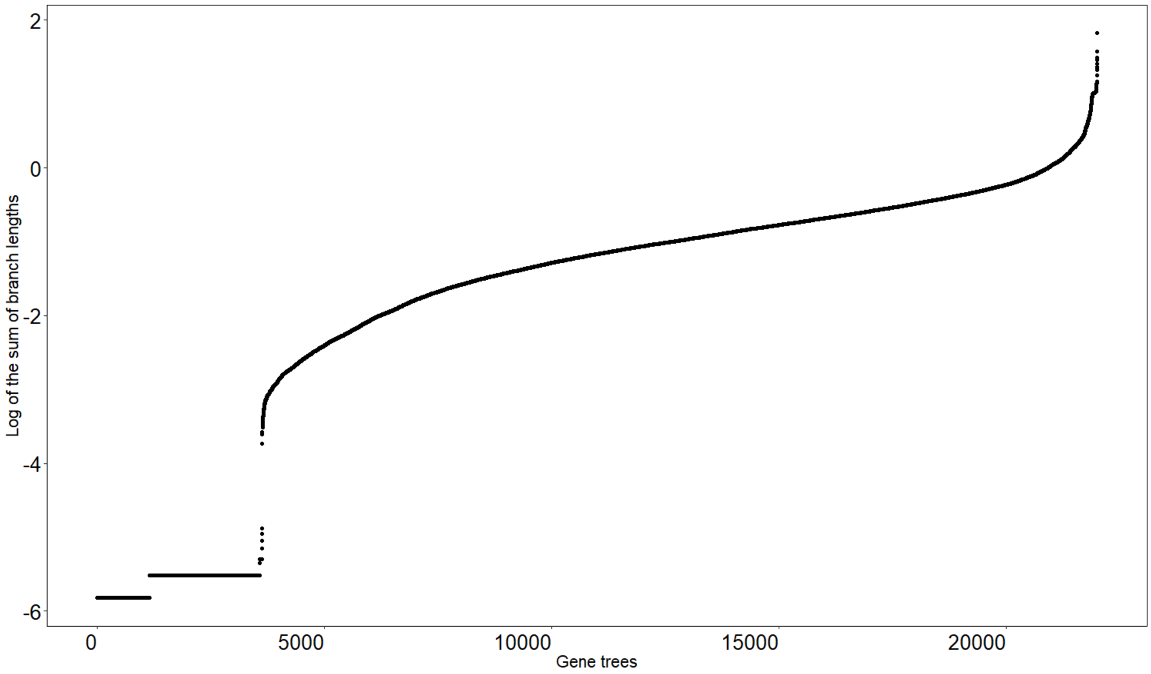


Figure 3.9: Dot plot showing the log of the sum of the branch lengths of each gene tree with at least four tips from the collection of genes of the 515 *E. coli* dataset on the Y axis and the 10,178 gene trees on the X axis. The sum of the branch lengths represents the amount of evolutionary information within one gene tree, where more evolutionary information equates to longer branch lengths and therefore a higher value for the sum of the branch lengths.

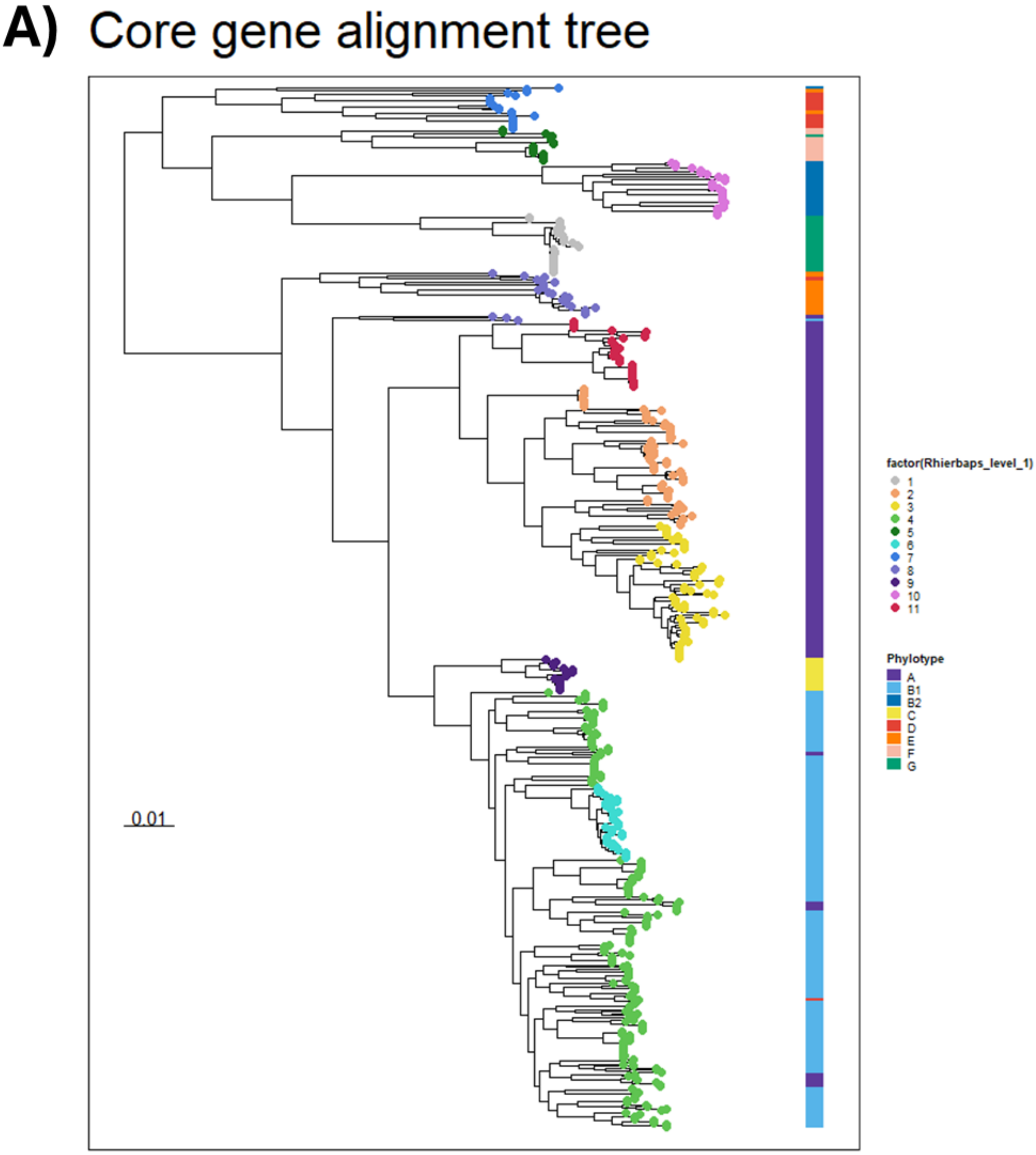
3.5.2 Selection of the standard approach for comparison

The research strategy for the most common methodology defined in the materials and methods identified 54 studies. Of which, 13 were excluded as eight studies investigated organisms other than *E. coli*, two were reviews, two focused on plasmids rather than the *E. coli* population, and one did not create a phylogenetic tree. Out of the 41 usable studies from this research strategy, the most commonly used approach and software was the core gene alignment core gene alignment approach using Roary (n = 15 studies) (Appendix 1).

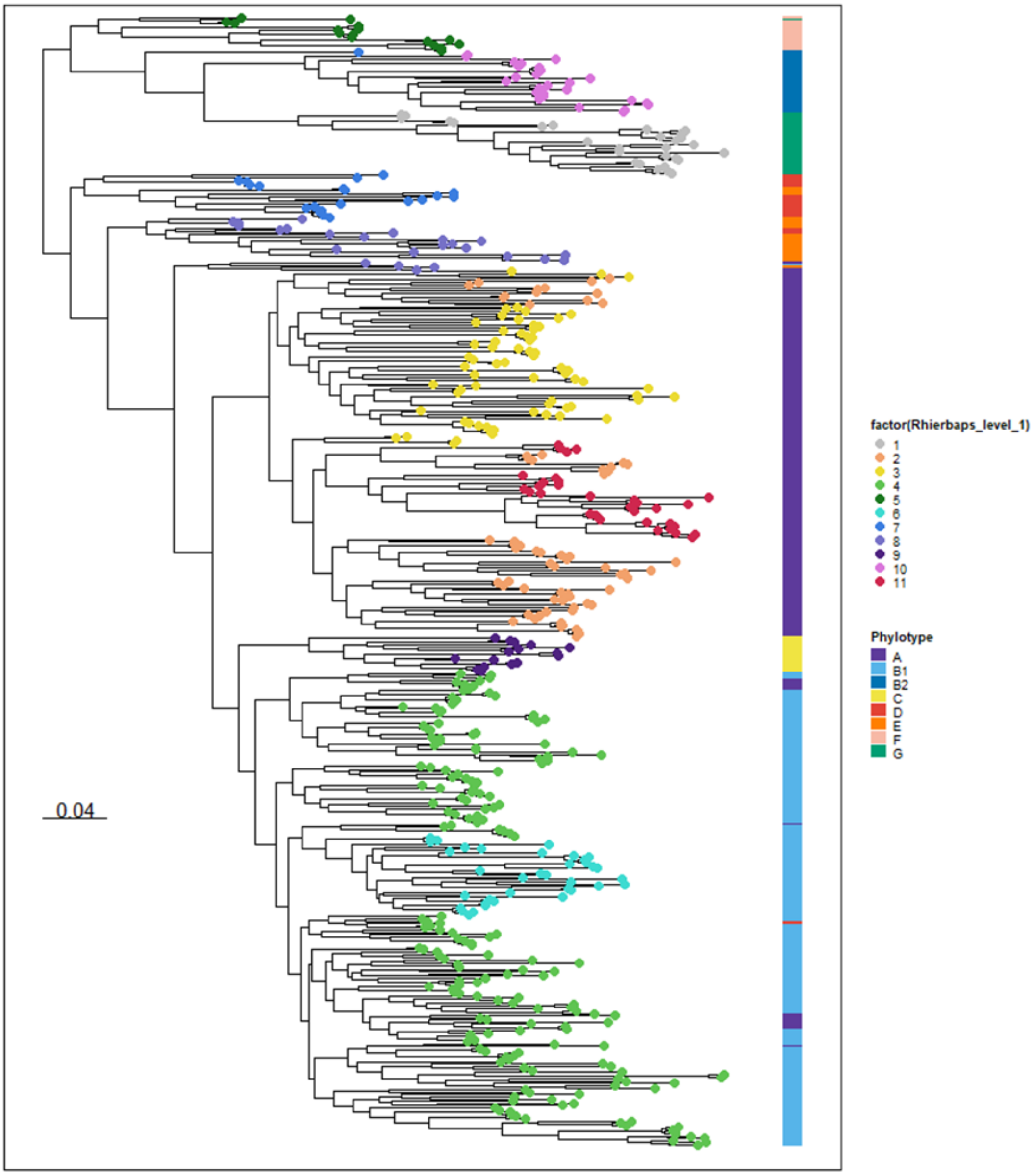
3.5.3 Comparison of phylogenomic approaches

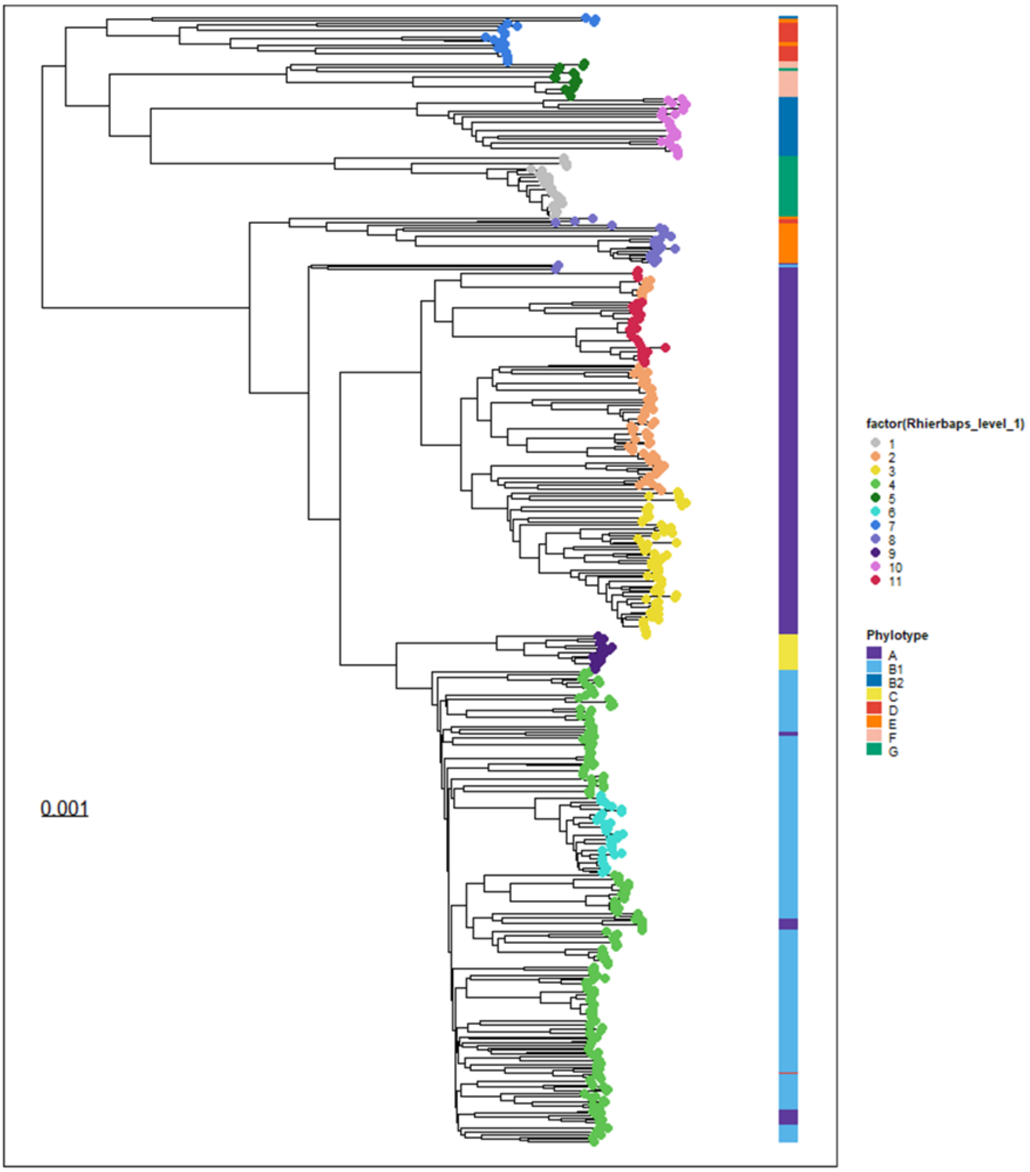
Phylogroup analysis revealed that phylogroups B1 (39.2%, n = 202) and A (35.5%, n = 183) were the predominant phylogroups in the 515 genomes. Phylogroups G (5.6%, n = 29), B2 (5.4%, n = 28) E (4.5%, n = 23), D (3.7%, n = 19), C (3.1%, n = 16) and F (2.9%, n = 15) were all found in lower proportions in the *E. coli* food survey dataset.

From visual inspection, the core gene alignment method, gene presence and absence method, MASH method, Phylonium method and Supertree methods were all found to cluster genomes in the phylogroups that were previously defined in the ECOR collection (Figure 3.10).


A)
Core gene alignment tree
0.01
factor(Rhierbaps_level_1)
1
2
3
4
5
6
7
8
9
10
11
Phylotype
A
B1
B2
C
D
E
F
G

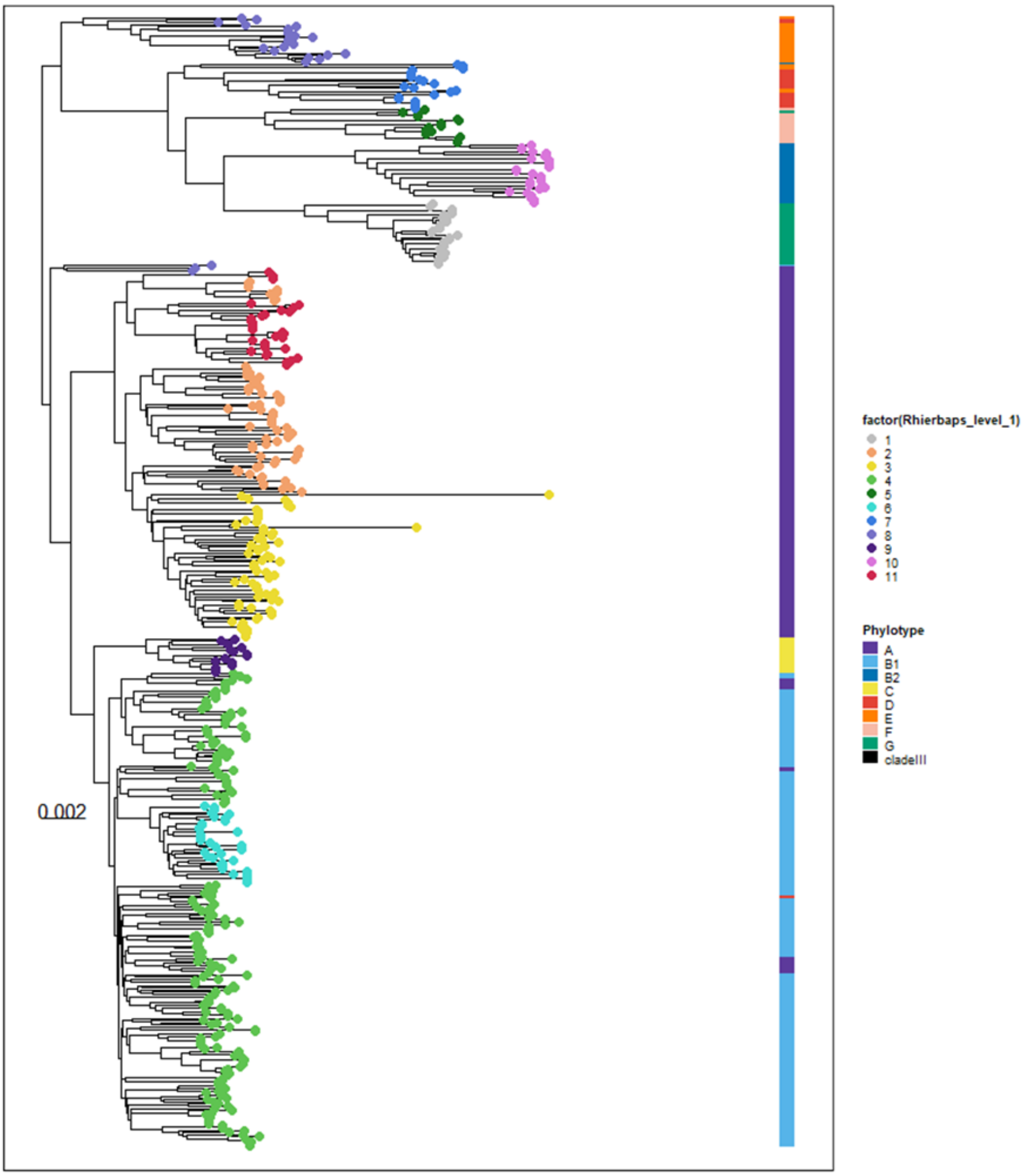
B) Gene presence and absence tree





C) Phylonium tree

D) MASH tree


0.002
factor(Rhierbaps_level_1)
1
2
3
4
5
6
7
8
9
10
11
Phylotype
A
B1
B2
C
D
E
F
G
cladeIII

E) Supertree

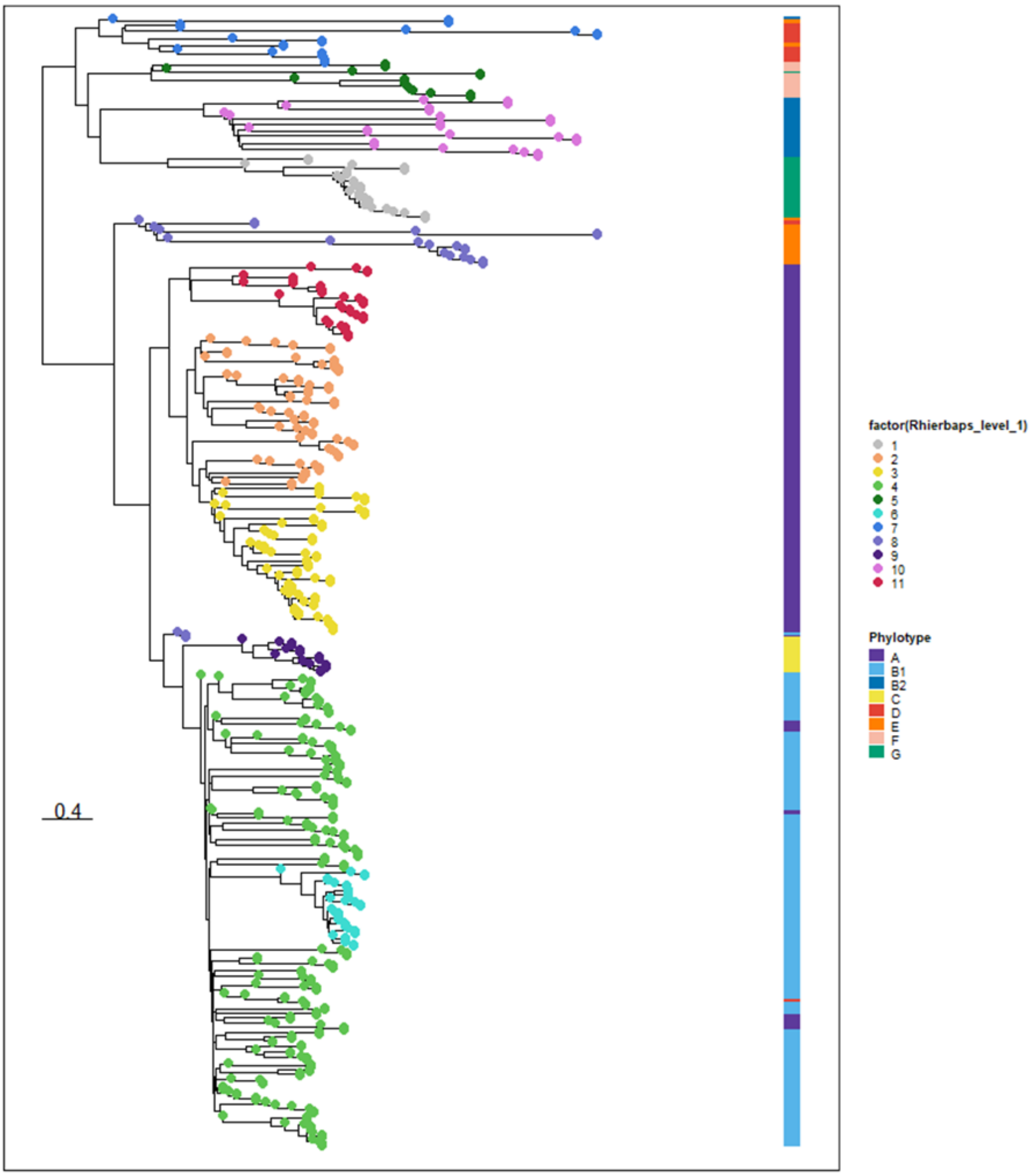

0.4
factor(Rhierbaps_level_1)
1
2
3
4
5
6
7
8
9
10
11
Phylotype
A
B1
B2
C
D
E
F
G

Figure 3.10: Phylogenetic trees representing 515 *E. coli* genomes from five different food types, clustered into 11 hierBAPS clusters derived from the core gene alignment – A) core gene alignment tree, B) gene presence and absence tree, C) alignment free Phylonium-derived tree, D) alignment free MASH-derived tree and E) Supertree. Phylogenetic distances for each resulting phylogenies are given by the scale bar which refers to the mean number of nucleotide substitutions per site for trees A, C, and D. The gene presence and absence tree scale bar represents the number of gene differences between organisms. The Supertree branch lengths are in coalescent units.

RhierBAPS used the core gene alignment to cluster the *E. coli* genomes into 11 population clusters termed “level 1”. Differences in the structure of the tree were visually observed. The core gene alignment tree was compared with the gene presence and absence tree. There was a difference was in the gene presence and absence tree where two *E. coli* isolates (EC18LG-0211-3 and EC18PR-0143-1) were moved into different phylogroups. Phylogroup D-BAPS cluster 7 was also placed away from phylogroup G-BAPS cluster 5 in the gene presence and absence tree compared to the core gene alignment tree. It was also found that phylogroup A-BAPS cluster 2 was distributed across the whole phylogroup and spread across BAPS cluster 11 and 3 (Figure 3.11).

The core gene alignment tree was also compared with the MASH tree. Phylogroup E-BAPS cluster 8 was moved next to phylogroup D-BAPS cluster 7 on the MASH tree. The topology of the tree indicates that this difference between the Supertree and MASH tree cannot be corrected by rotating the clades of the tree (Figure 3.11).

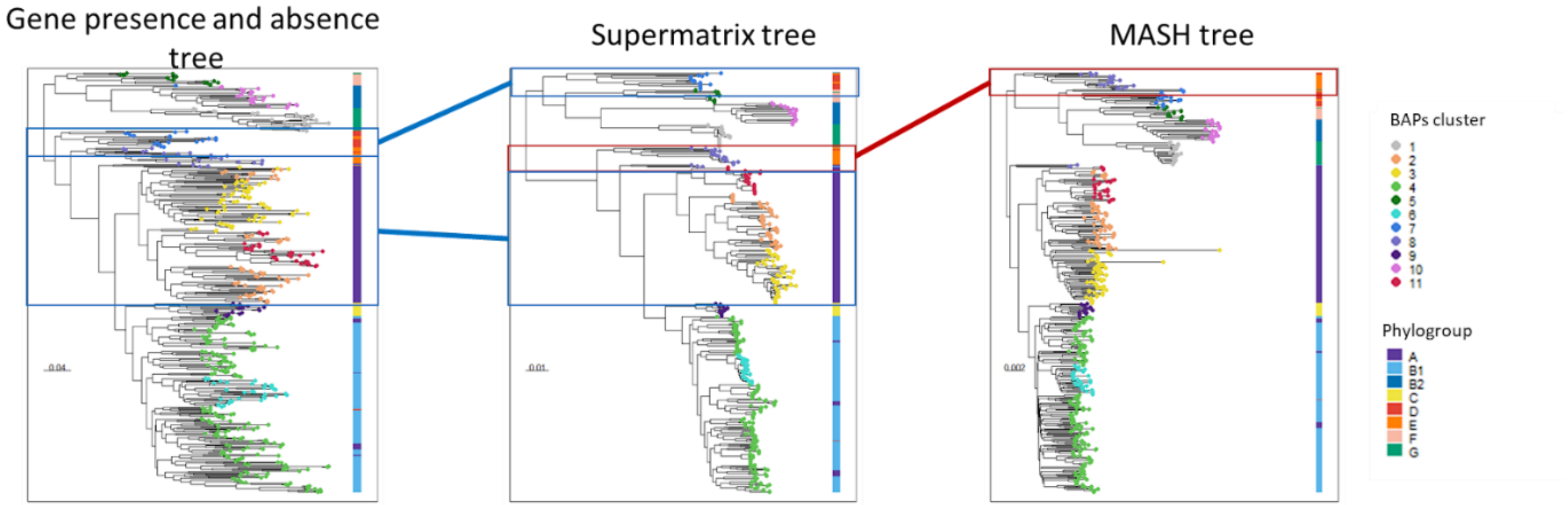


Figure 3.11: Visual comparison of topological differences between the core gene alignment **tree with the gene presence and absence tree and MASH tree representing 515 *E. coli* genomes from the food survey dataset. BAPS clusters are visualised as nodes on the end of the branch tips, and phylogroups are indicated by the colours on the matrix.**

To quantify the differences of the BAPS clusters onto the different topologies, the level 2 BAPS clusters were ordered to match the sequence of genomes for each tree. Silhouette scores of the BAPS clusters according to the tree tip order were calculated. The 3rd upper quantile was used to describe the top 25% of the silhouette width scores for each phylogenomic approach and the median was used to describe 50% of the silhouette scores for each phylogenomic approach. The BAPs population clusters were the best described by the core gene alignment tree (3rd upper quartile = 0.909; median silhouette score = 0.690), Phylonium tree (3rd upper quartile = 0.853; median silhouette score = 0.627) and Supertree (3rd upper quartile = 0.816; median silhouette score = 0.573) (Figure 3.12, table 3.2).

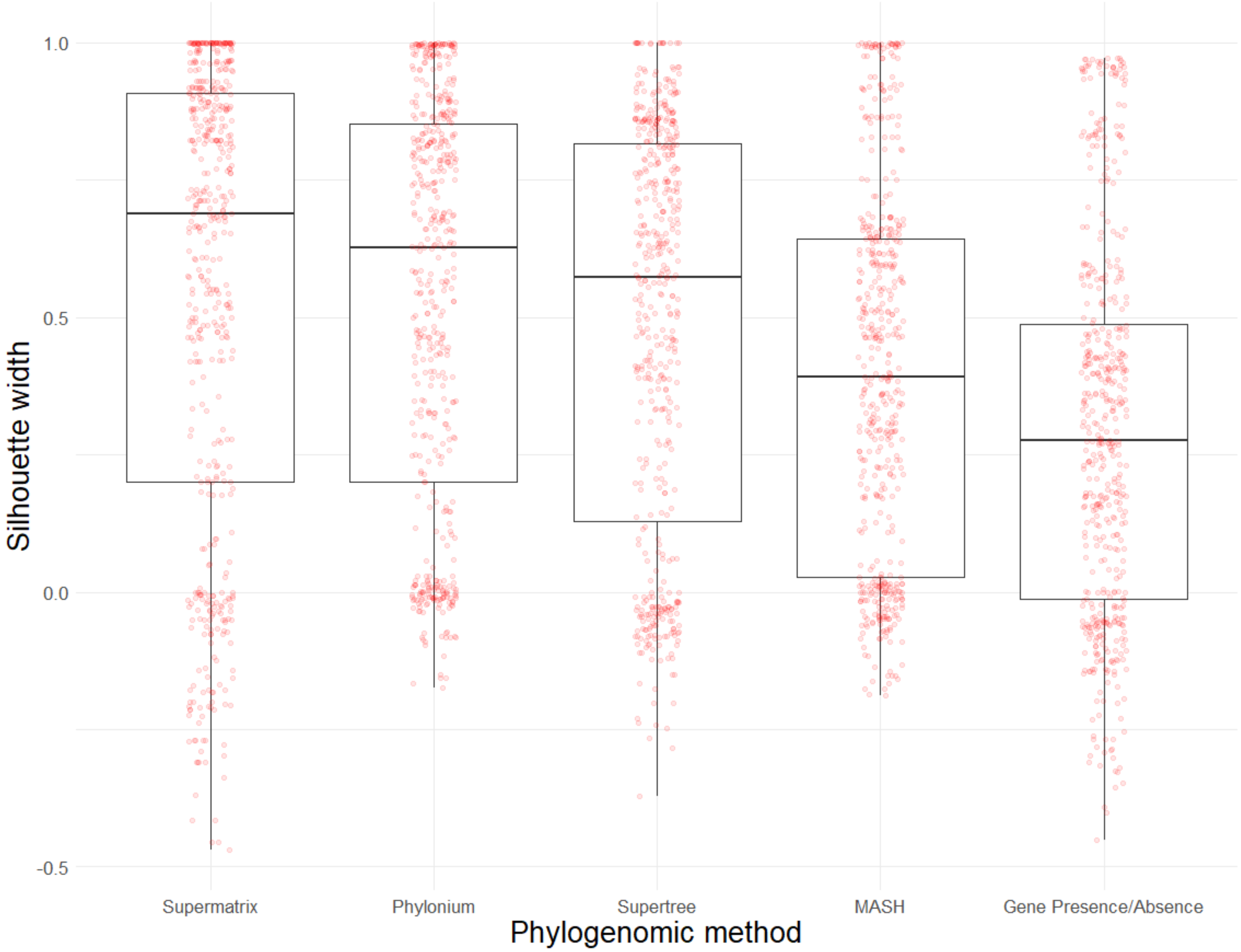


Figure 3.12: Comparison of silhouette scores to evaluate the measure of how well the BAPS clusters describe the clustering across the different phylogenomic methods.

Table 3.2: The 3rd quantile and median of the silhouette width indicating the cut-off for the top 25% and 50% of the silhouette width scores for each phylogenomic approach.

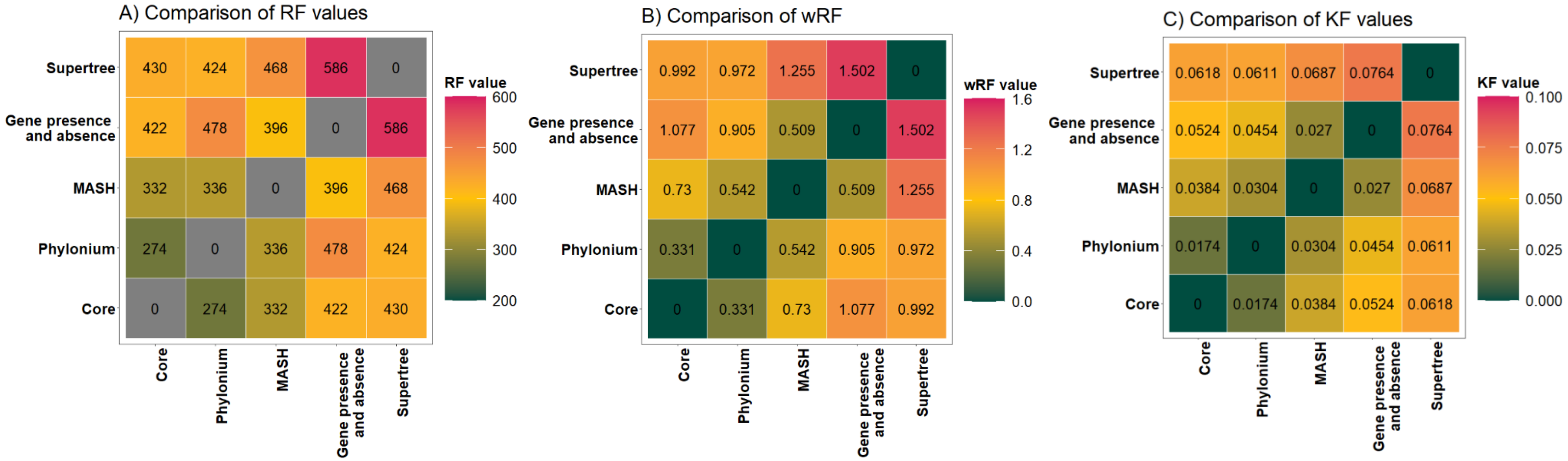
Phylogenomic approach	3rd Quantile of the silhouette width scores	Median silhouette width (±SE)
Core gene alignment	0.909	0.690 (0.644-0.726)
Phylonium	0.853	0.627 (0.550-0.678)
Supertree	0.816	0.574 (0.511-0.630)
MASH	0.644	0.391 (0.347-0.464)
Gene presence and absence	0.488	0.277 (0.243-0.333)

3.5.4 Comparison of topologies

The Robinson-Foulds distances (RF) between trees were compared to quantify the topological differences. The core gene alignment tree and the Phylonium tree were found to have the lowest RF distance (274 RF distance), indicating that the core gene alignment tree and Phylonium tree were the most similar. The gene presence and absence tree and the Supertree had the highest RF distance (586 RF distance), indicating that the gene presence and absence tree and Supertree were the most dissimilar.

The Supertree method was the most different according to RF values when compared to the other methods. The core gene alignment phylogeny had the lowest average RF value compared to all the other phylogenomic methods (Figure 3.13A).

The weighted Robinson-Foulds (wRF) distance and Kuhner-Felsenstein (KF) distance (branch score) were used to take the branch lengths into account. Both the wRF and the KF value/branch score consider the differences in branch lengths. The Phylonium tree and the core gene alignment phylogeny had the lowest wRF value (0.331 wRF value). The Supertree method again showed the highest differences for wRF value compared to the other methods (Figure 3.13B). The KF distance showed similar trends to the RF and wRF, where Phylonium and the core gene alignment phylogeny had the lowest KF/branch score value (0.0174 KF/branch score value). The Supertree method, again showed the highest differences for wRF value compared to the other methods (Figure 3.13C).



A) Comparison of RF values

	Core	Phylonium	MASH	Gene presence and absence	Supertree
Supertree	430	424	468	586	0
Gene presence and absence	422	478	396	0	586
MASH	332	336	0	396	468
Phylonium	274	0	336	478	424
Core	0	274	332	422	430

B) Comparison of wRF

	Core	Phylonium	MASH	Gene presence and absence	Supertree
Supertree	0.992	0.972	1.255	1.502	0
Gene presence and absence	1.077	0.905	0.509	0	1.502
MASH	0.73	0.542	0	0.509	1.255
Phylonium	0.331	0	0.542	0.905	0.972
Core	0	0.331	0.73	1.077	0.992

C) Comparison of KF values

	Core	Phylonium	MASH	Gene presence and absence	Supertree
Supertree	0.0618	0.0611	0.0687	0.0764	0
Gene presence and absence	0.0524	0.0454	0.027	0	0.0764
MASH	0.0384	0.0304	0	0.027	0.0687
Phylonium	0.0174	0	0.0304	0.0454	0.0611
Core	0	0.0174	0.0384	0.0524	0.0618

Figure 3.13: A) Comparison of the Robinson-Foulds distance (RF), B) weighted Robinson-Foulds (wRF) distance and C) Kuhner-Felsenstein (KF) distance/branch score between the five different phylogenomic methods using the normalised branch lengths from the core genome alignment (core), Phylonium, MASH, Gene presence and absence and Supertree phylogenies.

To summarise all previous results, the five phylogenomic approaches were compared to evaluate whether more than just the core genes were included in the analysis; whether phylogroups clustered as expected on the tree; how well the core gene alignment derived BAPS clusters mapped onto the tree; and to evaluate the similarity in topologies (Table 3.3). All methods except the core gene alignment method utilised genes other than those defined as core genes. All methods clustered the majority of the phylogroups on the trees as expected. The median silhouette score (±SE) was highest for the core gene alignment (0.690 score (SE = 0.644-0.726)) and Phylonium trees (0.627 (SE = 0.550-0.678)). The approaches with the lowest median RF values was the core gene alignment (292 RF value), Phylonium (302 RF value) trees. The approach with the lowest median wRF and KF value was the Phylonium approach (wRF value = 0.550, KF value =0.0309).

Table 3.3: Comparison of phylogenomic methods

Method	Alignment method	Accessory genes Included in analysis?	Phylogroups clustering together on the tree?	BAPs cluster median silhouette score (±SE)	Mean RF value	Mean wRF value	Mean KF value
MASH + MASHtree	Alignment free	Yes	Yes	0.391 (0.347-0.464)	306	0.607	0.0329
Phylonium + rapidNJ	Alignment free	Yes	Yes	0.627 (0.550-0.678)	302	0.550	0.0309

Table 3.3: Comparison of phylogenomic methods

Method	Alignment method	Accessory genes Included in analysis?	Phylogroups clustering together on the tree?	BAPs cluster median silhouette score (±SE)	Mean RF value	Mean wRF value	Mean KF value
Gene presence and absence matrix + rapidNJ	Alignment free	Yes	Yes	0.277 (0.243-0.333)	376	0.799	0.0402
Supertree (Roary + IQtree + Astral-Pro)	Gene trees aligned + quartets	Yes	Yes	0.574 (0.511-0.630)	382	0.944	0.0536

Table 3.3: Comparison of phylogenomic methods

Method	Alignment method	Accessory genes Included in analysis?	Phylogroups clustering together on the tree?	BAPs cluster median silhouette score (±SE)	Mean RF value	Mean wRF value	Mean KF value
core gene alignment + IQtree	Core gene alignment	No	Yes	0.690 (0.644-0.726)	292	0.626	0.0340

3.6 Discussion

This study hypothesised that the different phylogenomic methods - Core gene alignment, Gene presence and absence, MASH, Phylonium and Supertree - would produce different population structures of *E. coli* from the same dataset. The main aim of this chapter was to determine which phylogenomic method to take forward and apply to future chapters. Each phylogenomic method is a valid way of evaluating the *E. coli* population structure. No biological question was being answered in this chapter, but the use of various phylogenomic approaches can be justified depending on the biological question.

The trees constructed from these phylogenomic approaches indicate an overall repeating phylogroup and topological structure that was recreated regardless of the approach. All methods correctly clustered the majority of genomes from the same phylogroups together, suggesting a level of credibility for use of all methods for determining the population structure of the *E. coli* in this dataset. However, the way these approaches use the same genomic information differs. This can lead to smaller changes in tree topology and in inference of the population structure. For example, there may be evidence of recombination and horizontal gene transfer occurring in this dataset, as evident by the elongated branch lengths in the MASH tree. Alternatively, this may just be due to the MASH software and MASHtree software failing to interpret the data, as the other methods did not pick this up. This elongation of the branches did not affect the overall phylogroup classification but does affect the inferred evolutionary history between neighbouring genomes.

3.6.1 The Phylonium tree and the Supertree describe the core gene alignment BAPS clusters better than the MASH tree and Gene presence and absence tree.

The silhouette scores were obtained from mapping the core gene alignment BAPS clusters onto the trees created from the five different approaches. This quantified how well the tree described the BAPS clusters and provided clear evidence of differences in how the populations were structured. The core gene alignment method was determined to be the most popular recent approach for investigating

E. coli populations. Therefore, it was used as the standard for comparison. It is therefore unsurprising that, as the BAPS clusters were based on the core gene alignment that is used as input for the tree inference, the core gene alignment approach would have the highest silhouette scores across all methods. However, it is surprising that the core gene alignment approach is not technically better than all the others, as the standard error for the median silhouette score of the Phylonium tree overlapped with the core gene alignment tree.

Analysis revealed that the Phylonium tree and Supertree described the BAPS clusters better than the MASH tree and Gene presence and absence tree. This also coincides with the differences in the positions of the phylogroups and the order by which they branch. The phylogroup order in the core gene alignment tree, Supertree, and Phylonium trees are the same, whereas phylogroups F and D had changed positions in the Gene presence and absence tree, and we see phylogroups D and E change positions in the MASH tree. This indicates that the approach used can change how the evolutionary relationships are inferred downstream.

It is evident that inference of phylogenetic relationships at the isolate level will differ. While the Phylonium and Supertree approaches both displayed the highest correspondence with the core gene alignment BAPS clusters and phylogroups, the Phylonium approach described the BAPS cluster better than the Supertree. The Phylonium and Supertree approaches may present as useful alternatives to the core gene alignment approach as they have the highest silhouette scores. However, this does rely on the assumption that the core gene alignment is the most appropriate approach to use as the standard for comparison.

In any study and in this dataset, the true evolutionary history can never be found. This is because the contamination of *E. coli* on the food, the growth over time, and the evolution of the *E. coli* were never observed or measured. Additionally, there are a lack of studies investigating what the best practice is between core gene alignment-based approaches, Supertree-based approaches, or feature-based approaches (alignment-free and gene presence and absence approaches). The core gene alignment approach is the most commonly used current phylogenomic approach for investigating *E. coli* population genomics (Appendix 1), but there have also been studies using alignment free approaches on *E. coli* datasets (Abram *et al.*, 2021, Zielezinski *et al.*, 2017). Additionally, the gene presence and absence approach has been used to explore the accessory genes within *E. coli*, usually in tandem with another phylogenomic approach (Hall *et al.*, 2021, Mahazu *et al.*, 2021). The supertree has been used in previous studies (Skippington and Ragan, 2012, Chen *et al.*, 2018), but is not as commonplace as the other approaches.

3.6.2 Horizontal gene transfer may explain the differences between trees.

E. coli has been identified to have an open pangenome with varied accessory genes (Decano and Downing, 2019). In this dataset, 92.9% of genes were defined as non-conserved genes in the pangenome, some of which may have arisen from horizontal gene transfer. There has been disagreement on whether or not this horizontal gene transfer can obscure the phylogenetic relationship between genomes. It has been previously argued that the phylogenetic relationship between prokaryotes cannot be truly deciphered by a tree because of horizontal gene transfer (Doolittle, 1999). However, other studies contradict this by finding there is a strong phylogenetic signal in bacteria despite the horizontal gene transfer. For example, high levels of horizontal gene transfer above the normally observed range in *E. coli* start to affect the branch lengths on a core gene alignment tree (Touchon *et al.*, 2009). Touchon *et al*., further argues that the horizontal gene transfer occurs in certain genome locations termed “hotspots”, which does not interfere greatly with the phylogenetic signal.

The phylogroups remained relatively similar and there was minimal or no dispersion across the tree. Similarly, the BAPS clusters across the trees showed a consistent pattern across methods and clusters that were not spread across the trees. The amount of horizontal gene transfer in this dataset is unknown, but horizontal gene transfer is likely to have occurred as there are over 35,000 non-conserved genes. Nonetheless, regardless of whether horizontal gene transfer is pervasive in this dataset or not, it did not greatly affect the recurring broad population structure found by all methods. The placement of individual *E. coli* isolates across the tree did differ according to the approach used, which may be attributed to horizontal gene transfer. Alternatively, the differences in tree construction algorithms and evolutionary models for homologous sites may also be a factor that leads to differences across the various phylogenomic approaches in the placement of the *E. coli* isolates on the tree.

3.6.3 Different tree inference approaches were used but still produced similar phylogenies.

The five phylogenomic approaches used two different types of tree inference software. The whole genome feature-based approaches all produced a distance matrix which were used as inputs for a neighbour-joining software (Simonsen *et al.*, 2008, Katz *et al.*, 2019), and the sequence-based phylogenomic approaches both used the same maximum-likelihood software IQ-TREE (Nguyen *et al.*, 2015).

These tree inference approaches are fundamentally different. The neighbour-joining approach is a hierarchical clustering algorithm which creates the tree by creating two clusters, joining the clusters together with the aim to minimise the total sum branch lengths (Nguyen *et al.*, 2015). In contrast to this, the maximum-likelihood approach infers the tree topology by using nucleotide substitution models, which are probabilistic models that describe changes between homologous sites, to calculate the most probable and statistically significant connection between organisms (Felsenstein, 1981). It has been argued that maximum-likelihood methods are superior to other distance-based methods, as the neighbour-joining approach assumes that the homologous sites evolve at the same rate (Hasegawa *et al.*, 1991). Despite the Phylonium tree being created by a neighbour-joining algorithm, it produced a tree that had the most similar topology to the maximum-likelihood core gene alignment approach. This suggests that Phylonium is a good alternative approach, regardless of being ultimately a distance-based approach.

3.6.4 The MASH approach and Gene presence and absence approach may not be the most reliable for inferring the evolutionary history of this dataset.

The MASH method provides a measure of similarity between genomes by using k-mers of whole genomes and consequently, can provide insight into the differences and similarities between genetic content. Within this study, it can be observed that two genomes within Phylogroup A cluster 3 in the MASH tree displayed longer branches than the other methods. One explanation for this might be that these particular lineages experience accelerated molecular evolution, however, this would also then be observed in the other trees. Alternatively, this might be explained by an effect whereby MASH was found to be affected by random regions, leading to increases in distances within the tree (Röhling *et al.*, 2020). Additionally, It has been suggested that a MASH-based tree that is affected by horizontal gene transfer still offers a high general accuracy on simulated datasets (Zielezinski *et al.*, 2017).

The gene presence and absence tree will be affected by mobile genetic elements, like plasmids, and horizontal gene transfer. Genes that are acquired from horizontal gene transfer and are also not present in other genomes will cause the genomes to be inferred as more different and lead to longer branch lengths. This method provides an overview of the pangenome and has been argued to provide a better resolution of *E. coli* outbreak strains by taking the accessory genes into account (Kaas *et al.*, 2012).

The gene presence and absence method described the core gene alignment BAPS clusters the worst out of all the methods. This indicates that the *E. coli* that come from the same BAPS cluster do not necessarily have the same accessory genes present. However, the Gene presence and absence tree still recreated the phylogroup structure. This indicates that *E. coli* from the same phylogroups display similar enough gene content to cluster together. This finding is supported by previous work where the *E. coli* phylogroups display distinct enough protein families across the pangenome to segregate and group them (Abram *et al.*, 2021). The Gene presence absence tree is not the most reliable for inferring whether genomes come from the same ancestor. This method provides a greater resolution at a pangenome level, but at the cost of lower resolution for inferring the evolutionary relationship between genomes.

3.6.5 Comparison of the Core gene alignment, Supertree and Phylonium methods

Previous work has argued that the effect of horizontal gene transfer should not affect core genes and inference of the evolutionary history (Brochier *et al.*, 2002, Daubin *et al.*, 2003). Indeed, *E. coli* has been argued to accumulate more mutations through horizontal gene transfer than through vertical gene transfer (Walk *et al.*, 2007) and it has also been argued that, despite this high level of gene transfer, the signal of the inferred phylogenetic history is still strong (Touchon *et al.*, 2009).

The Supertree considers the evolutionary history of each gene including the accessory genes. The Supertree takes horizontally acquired genes into account when inferring a species tree. The 22,008 gene tree tips ranged from 4 – 2613 single-copy and multiple copy genes (median = 14), suggesting that the average gene is present in a few genomes and was likely acquired though HGT. The sum of the branch lengths for 22,008 single-copy and multi-copy gene trees indicates the maximum amount of information that can be used for the Supertree. The range of the log-transformed sum branch lengths were between -5.82 – 1.82 (median = -1.17) also shows that the average gene tree contained minimal amounts of information for inferring the final species tree.

The Supertree method is theoretically robust to changes created by horizontal gene transfer. Previous *E. coli* population studies have used the Supertree method (Chen *et al.*, 2018) and have also been found to reconstruct the phylogroups (Skippington and Ragan, 2012), albeit on a smaller scale. The quartet-based phylogeny (Supertree method) has also be argued to depict a strong phylogenetic signal in prokaryotes, despite horizontal gene transfer (Avni and Snir, 2020).

The effects of horizontal gene transfer on the core gene alignment tree and Supertree approach have been explored and previous work has shown that the core gene alignment approach is less likely to recover the correct tree topology when HGT affects more than 25% of the genes (Williams *et al.*, 2011). It was also found in the same paper that the correct topology was still obtained when around 45% of the genes underwent horizontal gene transfer using the Supertree method

(Williams *et al.*, 2011). This is further supported by previous Supertree work using simulated data with more than minimal rates of horizontal gene transfer (Lapierre *et al.*, 2014). In real datasets, the true rates of horizontal gene transfer are difficult to estimate, however in this dataset there is strong evidence showing that a large amount of the genes in the dataset were gained from horizontal gene transfer, meaning that the Supertree would likely be a better approach to use in theory.

Practically, the core gene alignment method is a straightforward approach that is less computationally intensive and faster. The core gene alignment approach also provides an estimate of nucleotide substitutions between genomes and has easier to understand bootstrap values than the Supertree (Lang *et al.*, 2013). The Supertree approach requires programming and creating scripts for individual steps, such as manual removal of certain labels on tips, and removal of gene trees with less than four tips. The Supertree approach also took 37 days to generate the gene trees from these 515 *E. coli* isolates. This would not be useful or appropriate in a disease outbreak scenario where a rapid response is needed, or in larger datasets with 1000 or more genomes. This approach is computationally intensive and slower than the other methods, which needs to be taken into account when choosing an approach for creating a phylogenomic tree.

Phylonium trees, on the other hand, are created by approximating a multiple sequence alignment from exact k-mer matches. Phylonium detects exact k-mer sequences termed “anchors” between compared genomes and evaluates the number of nucleotide mismatches between anchors. Phylonium terms this as an “approximate multiple sequence alignment”, which the software aims to make as large as possible (Klotzl and Haubold, 2020). This means that regions of the genome that extend past the strict 95% - 99% cutoff are included in analysis. Phylonium ignores non-homologous regions and is therefore theoretically robust against horizontal gene transfer. Additionally, this chapter did not directly measure the time taken for each approach, however, the Supertree and core gene alignment approaches had a greater number of intermediate steps to generate the final species tree than Phylonium. Phylonium is magnitudes faster than the Supertree and core gene alignment approach and even easier to use than the core gene alignment method.

In this chapter, Phylonium was one of the five methods used to evaluate the *E. coli* population in our dataset. To our knowledge, Phylonium has not yet been used to show the *E. coli* phylogenetic groups. It was therefore useful to see that this software did reconstruct the same population structure, as well as indicate the closest similarity to the core gene alignment. Additionally, the Phylonium tree had the lowest average RF, wRF and KF values when compared to all the other trees. This indicates that, at least for this dataset, Phylonium was the best alternative to creating the most similar tree out of these tested methods. Phylonium and the core gene alignment methods may create the most similar phylogenetic trees as both methods are based on the overall shared similarities between genomes (Table 3.3).

3.7 Conclusion

This chapter aimed to address three questions:

1) Will the Core gene alignment, Gene presence and absence approach, alignment free approaches and the Supertree approach show different population structures?

All approaches showed at least some minor differences in the distribution of BAPS clusters along the tips of the trees. However, the alignment-free approaches both indicated the same population structure and were also found to be the second and third most similar to the other approaches. The phylogroups were all found to be consistent, indicating that all methods are valid for clustering genomes in their phylogenetic groups, as defined by previous literature. When mapping the core gene alignment BAPS clusters onto the other trees, Phylonium and the Supertree were found to describe the clusters well. The differences in silhouette scores indicate that the order of the isolates changes depending on the approach used and therefore could affect the inference of the evolutionary relationship between isolates in a population.

2) Do the different approaches create different topologies based on the same population dataset?

As evident by the average RF, wRF and KF values, all methods created different topologies. This was also visually seen in the branch lengths and structure (Figures 3.6 and 3.7). When these RF, wRF and KF values were compared, the core gene alignment tree was the most similar to all other approaches. On the other hand, the Supertree method was the most dissimilar to all other approaches.

3) Identify and justify which approach would be the best suited for describing what types of *E. coli* are present from different sources.

The results from this chapter suggest that all approaches can infer the broader phylogenetic signal of *E. coli* in a population. There was no substantial difference in phylogenetic reconstruction of the other methods compared to the core gene alignment approach, however there are some advantages for using the other approaches. To reiterate, this is not necessarily the best choice, rather the most popular. This core gene alignment approach does not consider the accessory genes, meaning that information that may allude to how specific lineages survive on different sources is lost. The other methods can be more useful for inferring evolutionary processes like horizontal gene transfer which may be more useful for indicating how a species is more suited for a certain environment. Therefore, it would be best to opt for a method that takes more than a strict set of core genes into account when determining the phylogenetic similarity between genomes (Table 3.3).

The Supertree and Phylonium approaches were the most similar to the core gene alignment method according to the BAPS clusters and silhouette scores. Phylonium created the tree that was the most like the core gene alignment tree and had the added benefit of taking more than just the core genes into account when constructing the tree. Additionally, Phylonium was found to be the most similar to all other methods, suggesting that Phylonium is a good middle ground for these five approaches (Table 3.3).

Therefore, in the next chapters, Phylonium will be applied to different *E. coli* population datasets to explore the types of *E. coli* present in retail foods, dogs, and humans.

Chapter 4: Whole genome sequencing multi-isolate approach for investigating *E. coli* diversity and contamination on retail foods from Norfolk, United Kingdom

Chapter contributions: Design of the food survey and sample collection was done by Mather, A.E., Janecko, N., Bloomfield S.J., and Palau, R. Collection of the food samples, as well as growth, detection and isolation of the *E. coli* were done by Bloomfield S.J., and Palau, R. Data analysis, scripts and code were written by Astorga, G. with help from Zamudio, R., and de Oliveira Martins, L. Astorga, G., Mather, A.E., Singh, D., Wain, J., Zamudio, R., and de Oliveira Martins, L. all discussed the work.

Publication: This chapter has been accepted for publication in Microbial Genomics under the title "A multi-isolate genomic approach identifies diverse *Escherichia coli* contamination and antimicrobial resistance carriage on retail foods"

4.1 Introduction

Food safety is regulated in the UK by the Food Standards Agency (FSA) and the UK Health Security Agency (UKHSA). One of the FSA key aims is to safeguard the general public health by reducing the likelihood of foodborne illnesses (Food Standards Agency, 2024, Simonsen *et al.*, 2008). Retail foods, when contaminated with an unsafe level of microorganisms (Janecko *et al.*, 2023), can suggest a lack of safety precautions or hygienic practices during food manufacturing and production (Health Protection Agency, 2009). The consumption of contaminated, improperly cooked or improperly washed food items may lead to gastrointestinal disease (Food Standards Agency, 2024).

Specific groups of microorganisms called indicator organisms are used to measure the amount of contamination on food. An example of a commonly used group of indicator organisms are coliform bacteria which are bacteria commonly found in soil, water, and the guts of humans and animals (Kay and Fricker, 1997). Coliform can include genera like *Escherichia, Citrobacter, Enterobacter,* and *Klebsiella* (Halkman and Halkman, 2014). However, using total coliform counts has been shown to lead to erroneous conclusions for food safety, as the presence of coliforms may be from environmental contamination and not necessarily faecal origin (Santos MI *et al.*, 2012). Microbiological criteria for specific microorganisms were defined for food safety measures in 2005 and are periodically updated (European Commission, 2005). Food safety measures currently use total aerobic bacteria counts, total *Enterobacteriaceae* counts, or look for specific organisms like pathogenic lineages of *E. coli* (European Commission, 2005, Health Protection Agency, 2009).

Typically, *E. coli* contamination is assessed by enumeration on food items via colony forming units (CFU) (UK Health Security Agency, 2023a) or most probable number (MPN) (International Organization for Standardization, 2015). Safe limits have been set for microbial contamination on meat carcasses, shellfish and vegetables in the UK (European Commission, 2005), but an estimated 2.4 million cases of foodborne illnesses were predicted to have occurred in the UK in 2018

(Holland and Mahmoudzadeh, 2020). In the event of an outbreak, the UK Health Security Agency (UKHSA) uses biochemical tests to identify *E. coli*, which can be further analysed using whole genome sequencing (WGS) (UK Health Security Agency, 2023b).

Short read WGS has been used by the FSA and UKHSA to primarily focus on pathogenic lineages and antimicrobial resistant lineages (Dallman *et al.*, 2013, Brouwer *et al.*, 2023, Day *et al.*, 2019). Additionally, studies that use WGS often do not make the number of sequenced *E. coli* per sample clear (Reid *et al.*, 2020, Randall *et al.*, 2021, Brouwer *et al.*, 2023), often taking one isolate as a representative per sample (Aworh *et al.*, 2024, Wang *et al.*, 2024, Liu *et al.*, 2024). As *E. coli* is known to be a genetically diverse organism (Geurtsen *et al.*, 2022), the selection for specific types of *E. coli* or taking one *E. coli* per sample will likely underestimate the genetic diversity and risk to consumers. Additionally, different food commodities will have different associated risks, which have not yet been thoroughly compared and assessed (Reid *et al.*, 2020, Randall *et al.*, 2021, Brouwer *et al.*, 2023, Aworh *et al.*, 2024, Wang *et al.*, 2024, Liu *et al.*, 2024).

4.2 Aims and Objectives

This study aimed to investigate the genomic diversity of *E. coli* on raw meats, leafy greens and seafood, which addresses three questions:

1) What types of *E. coli* contaminate retail foods in this collection of retail foods from Norfolk, United Kingdom?

2) How diverse are the populations of *E. coli* within food commodities, and within a sample?

3) What potential health risks do these contaminating *E. coli* present to consumers?

4.3 Materials and methods

4.3.1 Sample collection and culturing

A collection of 516 *E. coli* isolates were collected prior to this thesis and used for chapter 3. Additional *E. coli* positive samples where four isolates were taken per sample were selected for additional whole genome sequencing.

As briefly described in chapter 2, *E. coli* positive samples were selected from a study described by Janecko and Zamudio *et al.* (Janecko *et al.*, 2023). Janecko and Zamudio *et al*., conducted a cross-sectional study in Norfolk, United Kingdom between May 2018 to November 2019 to investigate pathogen contamination. The sampling focused on five food commodities that reflected highly consumed food products based on the Family Foods 2015 consumption data (Department for Environment Food and Rural Affairs, 2017). The five food commodities included raw chicken, raw pork, raw/cooked prawns, raw salmon, and pre-packaged leafy greens, resulting in 1,369 food samples (Janecko *et al.*, 2023).

Sample processing, preparation and *E. coli* microbiological detection were previously described in Janecko and Zamudio *et al.* (Janecko *et al.*, 2023) and was stated in Chapter 2 of this thesis (**See Section 2.1.2**).

The methods for isolation and biochemical confirmation were stated in Chapter 2 of this thesis (**See Section 2.1.2**). Each *E. coli*-positive sample had up to four *E. coli* colonies stored at −70 °C in 1 mL of Brucella broth +17.5% glycerol (ThermoFisher Diagnostics, Rochford, UK).

4.3.2 Initial datasets and sample selection

In the study from Janecko and Zamudio *et al.*, 859 out of the 1,369 samples were *E. coli* positive (Janecko *et al.*, 2023). Prior to the start of this project, 516 *E. coli* isolates from 197 samples were selected for WGS.

The samples for before this project were collected by Janecko, N., Bloomfield S.J., and Palau, R., and Mather, A.E., for four reasons and are stated in Chapter 2 of this thesis (**See Section 2.1.3**).

The diversity of *E. coli* contaminating retail foods was investigated and Astorga, G. prioritised samples with four isolates for further sequencing. In total, 401 *E. coli* positive samples with up to four *E. coli* isolates per sample were used in this study. The 401 samples included 126 raw chicken samples, 115 raw pork samples, 75 raw and cooked prawns, 52 raw leafy green samples and 33 raw salmon samples.

4.3.3 Whole genome sequencing, quality control and assembly

DNA extraction was described in chapter 2 (**See Section 2.3**). and additional DNA extractions for this chapter were performed by Bloomfield S.J., and Palau, R. Whole genome sequenced isolates are hereafter referred to as genomes. Quality control processes are described in chapter 2. The sequence data generated in this study are available in the Sequence Read Archive (SRA: PRJNA1177805 and SRA: PRJNA1107692) (Appendix 2). The data analysis, scripts and code were written by Astorga, G. with guidance from Zamudio, R., Bloomfield, S.J., de Oliveira Martins, L., Singh, D., Wain, J., and Mather, A.E.

4.3.4 *E. coli* phylogroup, sequence typing, and phylogenetic analysis

E. coli genome assemblies were classified into phylogroups and sequence types (STs) as stated in **Section 2.4.2.**

Unclassified STs were further characterised into matching novel STs by identifying unique combinations of *adk, fumC, gyrB, icd, mdh, purA,* and *recA* profiles provided by the MLST output (Appendix 3). Phylonium v.1.6 (Klotzl and Haubold, 2020) with default settings estimated the phylogenetic distances between the *E. coli* assemblies. The output from Phylonium v.1.6 was input into rapidNJ v.2.3.2 (Simonsen *et al.*, 2008) with the -i pd -o t -x options.

The phylogenetic tree was visualised with R v.4.1.3 (R Core Team, 2022) in RStudio v. 2021.09.1 (RStudio Team, 2020) using the ggtree v.3.2.1 (Yu, 2020), ape v.5.5 (Paradis and Schliep, 2019), phytools v.1.0-1(Revell, 2012), and ggnewscale v.0.4.5 (Campitelli, 2022) packages. Plots were visualised using ggplot2 v. 3.3.5 (Wickham, 2016). Clades enriched for a food commodity were identified during phylogenetic inspection and were selected for further analysis.

4.3.5 Pairwise SNP analysis

Reference genomes were chosen using Referenceseeker v. 1.8.0 (Schwengers *et al.*, 2020) and the bacteria RefSeq database (downloaded January 2024) (O'Leary *et al.*, 2016) on the collection of *E. coli* grouped into their respective STs. The reference that was chosen for the majority of the genomes (minimum > 50%) with the highest average nucleotide identity and highest conserved DNA was chosen as the reference for that ST.

Reads from each ST of interest were aligned to their respective reference genome (ST10 reference (Genbank Accession assembly ID: GCA_900636145.1), ST101 reference (Genbank Accession assembly ID: GCA_029318875.1), and ST117 reference (Genbank Accession assembly ID: GCA_025369795.1)) using Snippy v. 4.6.0 (Seemann, 2015). Gubbins v. 3.2.0 (Croucher *et al.*, 2014) was used to remove SNPs attributed to putative recombination from the alignments. The pairwise SNP differences were calculated using SNPdist v.0.7 (Seemann *et al.*, 2021). The threshold to determine the epidemiologically linked STs within the same sample was five SNPs, as this is the UKHSA threshold for an outbreak of Shiga toxin-producing *E. coli* O157 (Mikhail *et al.*, 2018).

4.3.6 Diversity between food commodities

The *E. coli* ST diversity (richness and inverse Simpson index) was estimated for each food commodity. The R package iNEXT v. 3.0.1 (Hsieh *et al.*, 2016) rarefied and extrapolated the data to predict and compare diversity between food commodities. Richness (q =0), and inverse Simpson index (q = 2) were calculated using the "incidence_freq" datatype, with confidence intervals set to 95% and the number of bootstraps set to 500.

4.3.7 Antimicrobial resistance genes, plasmid replicons and virulence genes

The presence of AMR genes, plasmid replicons and virulence genes were determined using ARIBA v.2.14.6 (Hunt *et al.*, 2017) with default settings on the trimmed reads. The Resfinder (Bortolaia *et al.*, 2020), PlasmidFinder (Carattoli and Hasman, 2020), and VFDB (virulence factor database) core (Chen *et al.*, 2015) databases were all downloaded in May 2023. Known point mutations associated with AMR were identified using PointFinder (Zankari *et al.*, 2017) and the PointFinder database (downloaded in October 2024) (Zankari *et al.*, 2017).

The output of ARIBA v.2.14.6 (Hunt *et al.*, 2017) with the Resfinder database (Bortolaia *et al.*, 2020) (downloaded May 2023) provided a gene name and an NCBI accession code, which allowed tracking of the associated sequence in the repository. The output of PointFinder (Zankari *et al.*, 2017) provided PubMed

Identifiers for papers that described known point mutations that confer AMR. The drug class for AMR genes and point mutations were determined by using the associated paper and cross-referencing the AMR gene name in comprehensive antibiotic resistance database (CARD) (McArthur *et al.*, 2013). The acquired AMR genes that may confer AMR and identified point mutations using the Resfinder and PointFinder database are collectively called AMR determinants.

The AMR determinants were classified by importance according to the World Health Organization scheme for critically important antibiotics (World Health Organization, 2019). Potential Multidrug-resistance (MDR) in this study was classified as the presence of AMR genes conferring resistance to three or more different antimicrobial classes according to the WHO antimicrobial class list (World Health Organization, 2019, Catalano *et al.*, 2022). AMR profiles were defined as the unique combination of AMR determinants for each genome, and expected phenotypes or intrinsic resistance within *E. coli* was not taken into account.

Typical genetic markers for diarrheagenic *E. coli* were investigated (Enteropathogenic *E. coli – eae, bfp*; Shiga-Toxin producing *E. coli*/Verocytotoxigenic *E. coli – stx1, stx2*; Enterotoxigenic – *elt, estA/estB;* Enteroinvasive *E. coli* – pINV; Enteroaggregative – pAA; Diffuse adhering *E. coli* – Afa/Dr family; Adherent-invasive *E. coli* – NA) (Geurtsen *et al.*, 2022). *E. coli* was also characterised as extraintestinal pathogenic (ExPEC) lineages if two or more of the following virulence genes were present in the genome: *papAH, papC, sfa/focDE, afa/draBC, iutA,* and *kpsMII* (Johnson *et al.*, 2003).

4.4 Results

4.4.1 Food samples

In this study, 401 food samples were collected from shops across Norfolk, United Kingdom. All chicken, all pork and all salmon samples were raw and all 52 leafy green samples were fresh.

4.4.2 *E. coli* contamination on food

A total of 1,067 *E. coli* isolates from chicken (n = 310 isolates), leafy greens (n = 162 isolates), pork (n = 279 isolates), prawns (n = 206 isolates) and salmon (n = 110 isolates) underwent whole genome sequencing and passed quality control (Table 4.1; Appendix 2).

Table 4.1: Number of samples and number of isolates per sample across each food commodity purchased in Norfolk, UK between May 2018 and November 2019.

Commodity	Number of samples with 1 isolate per sample	Number of samples with 2 isolates per sample	Number of samples with 3 isolates per sample	Number of samples with 4 isolates per sample	Total number of samples	Total number of isolates
Chicken	22	59	10	35	126	310
Leafy greens	7	9	7	29	52	162
Pork	16	63	7	29	115	279
Prawn	5	39	1	30	75	206
Salmon	2	7	2	22	33	110

Total samples	52	177	27	145	401	1,067

In silico MLST classification assigned the 1,067 *E. coli* genomes in this collection into 238 known STs and 54 unknown STs (Appendix 3). In this collection, the five most abundant STs comprised 19.5% of the genomes - ST10 (n = 69), ST117 (n = 39), ST101 (n = 35), ST155 (n = 26), and ST216/ST5474 (n = 23) (Figure 4.1A). Chicken (n = 118 STs) and pork (n = 103 STs) samples had the highest number of unique STs, followed by prawns (n = 75 STs), then leafy greens (n = 68 STs) and salmon samples had the lowest number of unique STs (n = 22 STs). Additionally, 73 STs were unique to chickens, 41 STs were unique to leafy greens, 55 STs were unique to pork, 47 STs were unique to prawn and 8 STs were unique to salmon.

There were 35 defined clonal complexes, however, the 63.4% of the genomes (n = 676) did not belong to a defined clonal complex. The most common clonal complexes were ST10 clonal complex (8.25%, n = 88), ST155 clonal complex (4.97%, n = 53), ST101 clonal complex (3.66%, n = 39), ST86 clonal complex (2.34%, n = 25), and ST23 clonal complex (2.06%, n = 22). Aside from ST155, the most prevalent STs were not present in all food commodities. ST10 was not found on salmon samples, ST117 was only found on chicken and pork samples, ST101 was not found on leafy green and salmon samples, ST216 was not found on chicken and leafy green samples and ST5474 was only found on salmon and pork samples (Figure 4.1B).

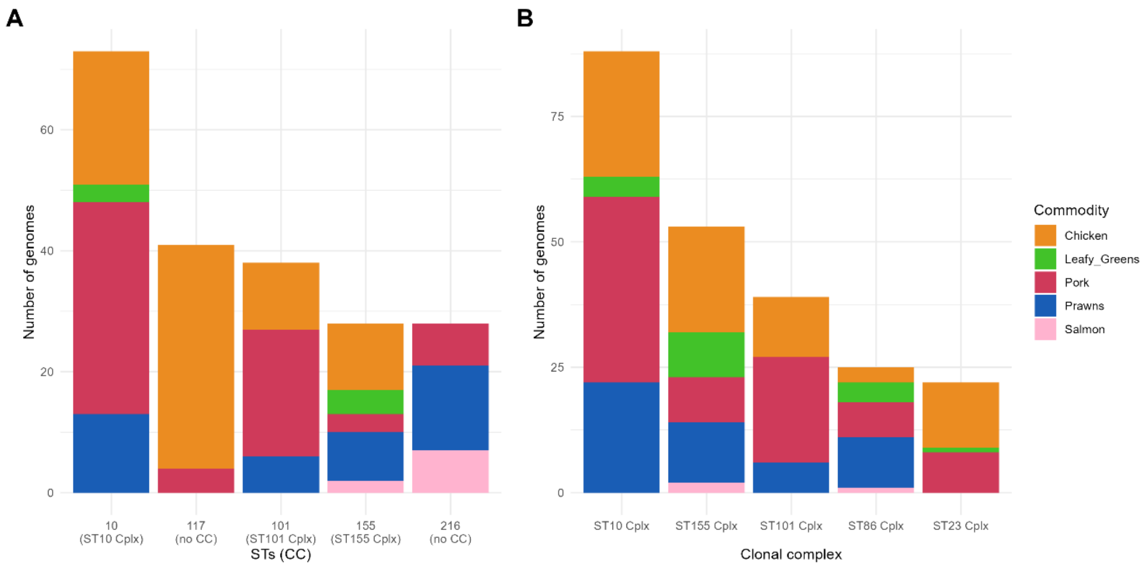


Figure 4.1: Stacked bar plot indicating (A) the number of *E. coli* genomes found for the most prevalent STs and (B) for the most prevalent clonal complexes in this dataset, categorised by food commodity.

4.4.3 Diversity estimates across commodities

The estimated ST richness appeared similar across all commodities except salmon, which had a smaller number of observed and estimated STs compared with the other commodities. The Simpson's diversity of leafy greens was higher than the other commodities, reflecting a more even ST distribution when compared to the other food commodities. Of the 68 distinct STs observed in 52 samples of leafy greens, only nine STs were observed in two different samples and one in three samples. However, in both pork and chicken there were STs that were common to many more samples, for example ST10 (n = 20) and ST101 (n = 14) being found in many of the 115 pork samples, and ST117 (n = 29) and ST10 (n= 18) seen in many of the 126 chicken samples (Figure 4.2).

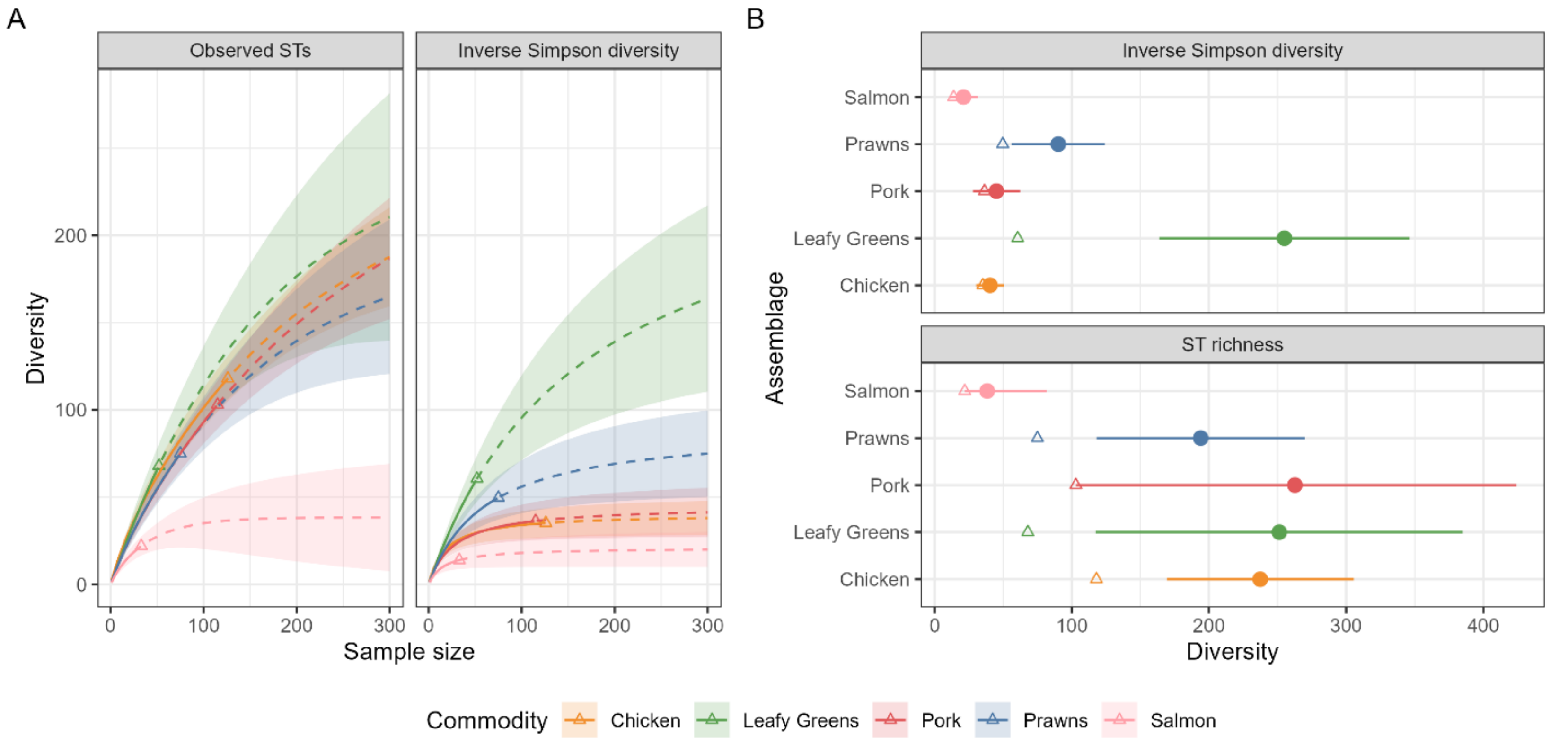


Figure 4.2: A) The rarefaction curves and the extrapolated diversity indices for the sequence types (STs) on each commodity with bootstrapped 95% confidence intervals. Points represent the observed number of distinct STs, and the number of samples collected. Solid lines are rarefaction curves, and dotted lines represent extrapolations. Shaded areas are 95% confidence intervals based on bootstrapped standard errors. B) The asymptotic diversity indices (dots with 95% confidence intervals, along with the observed diversity [triangles] for each commodity) coloured by food commodity.

4.4.4 *E. coli* sequence type diversity within samples

In this dataset, 48.1% of samples (n = 193 samples) had two or more STs, and 13.7% (n = 55) of samples were contaminated with three or more STs. Salmon samples were the only food commodity observed to be contaminated with a maximum of two STs (Figure 4.3).

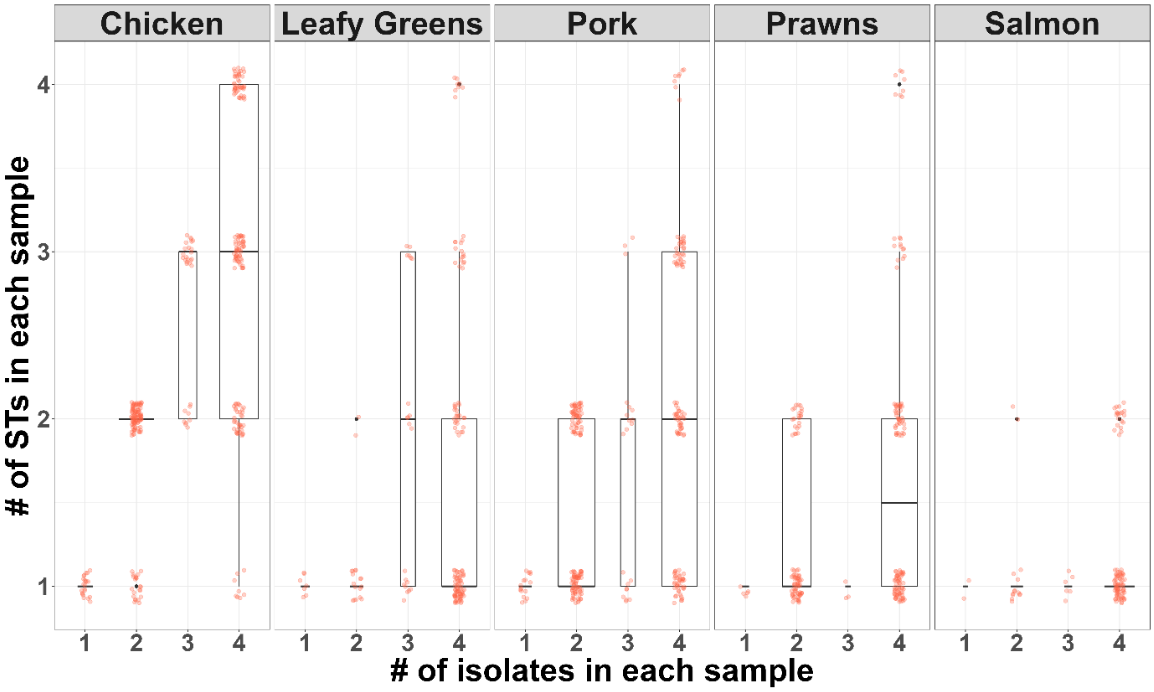


Figure 4.3: Boxplots of the number of sequence types (STs) in each of the 401 retail chicken, pork, prawns, salmon, and leafy greens samples. The Y axis shows the number of STs when up to four *E. coli* isolates in each sample were sequenced. Each red dot represents a sample. The box shows the interquartile range of the number of STs in each sample.

The three most abundant STs were ST10 (n = 73 genomes), ST117 (n = 41 genomes) and ST101 (n = 38 genomes). There were 34 samples which were contaminated with at least two ST10s (n = 18 samples), ST117s (n = 8 samples), or ST101s (n = 8 samples). There were 77 pairwise ST comparisons within the same sample from these three STs. In 22 cases, the SNP diferences between genomes within the same sample and ST exceeded five SNPs, with a maximum of 845 pairwise non-recombinant single nucleotide polymorphisms (SNPs) from two ST10 genomes from the same sample (Figure 4.4). Salmon was the only food commodity without multiple contaminating ST10s, ST101s, or ST117s on the same sample.

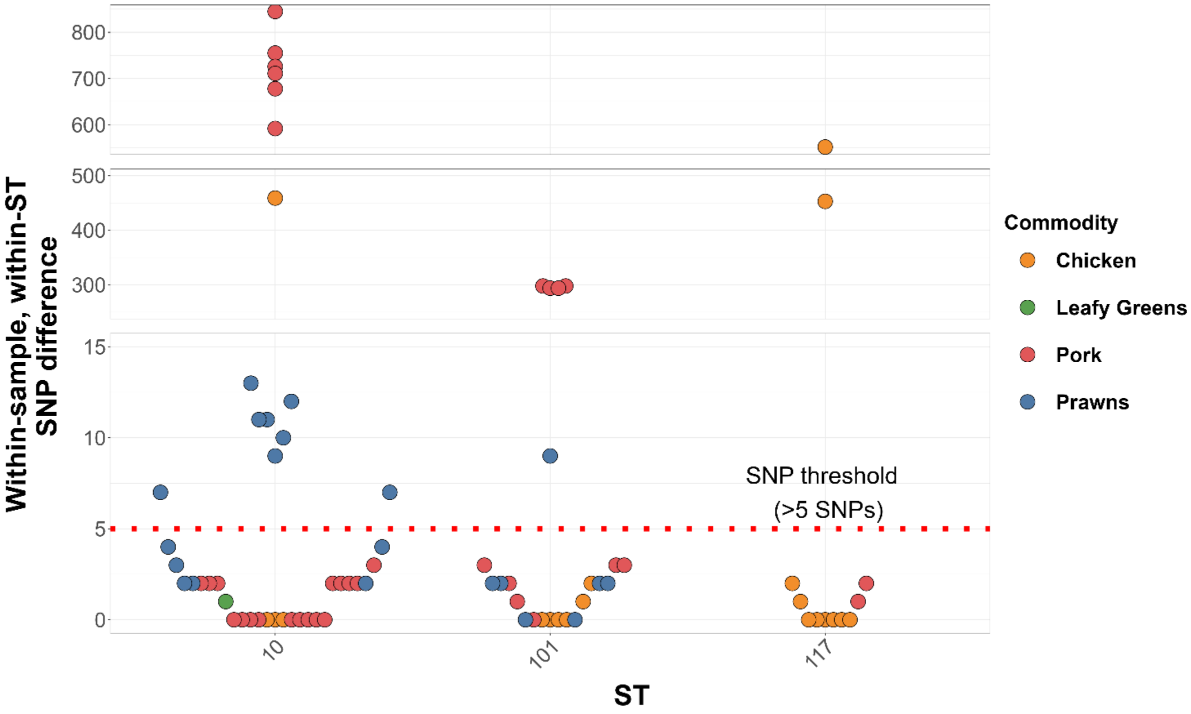


Figure 4.4: Comparisons of the within-sample within-sequence type (ST) single nucleotide polymorphism (SNP) differences across ST10, ST101 and ST117 within the collection of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens. Within-sample within-ST comparisons are coloured by the food commodity from which they were isolated. The five SNP threshold for epidemiologically linked *E. coli*, as stated by the United Kingdom Health Security Agency (UKHSA) Shiga toxin-producing *E. coli* O157 (Mikhail *et al.*, 2018), is indicated by the red dotted line.

4.4.5 *E. coli* plasmid replicons, virulence genes and antimicrobial resistance genes

In this dataset, 63 unique plasmid replicon types were identified, with the three most frequent belonging to the IncF incompatibility type. The median number of different plasmid replicons in each genome was 2 (range = 0-9).

The most frequent plasmid replicon was IncFIB_AP001918 which was found in 417 genomes (n samples = 204). Of these 204 samples with IncFIB_AP001918-harbouring *E. coli*,136 samples also harboured AMR determinants. The three most frequent AMR determinants that co-occurred with IncFIB_AP001918 were blaTEM-1b (n genomes = 133; n samples = 79), tet(A) (n genomes = 126; n samples = 79), and sul2 (n genomes = 115; n samples = 77).

IncFIB_AP001918 was observed in *E. coli* from all food commodities but was most frequently observed in 113/126 chicken samples. This was followed by pork (n samples = 55/115 samples), Leafy greens (n samples = 15/52), prawns (n samples = 17/75) and salmon (n samples = 4/33).

No typical intestinal pathogen (InPEC) genetic markers *stx1, stx2*, *eae, elt, estA/estB, bfp,* or Afa/Dr family were found in this dataset. The ExPEC virulence factor scheme identified 26 putative ExPEC across 10 chicken samples (genomes = 13) and seven pork samples (genomes = 13). Putative ExPEC belonged to eight known and two unknown STs; ST117 had the highest number of genomes classified as putative ExPEC (n samples = 7, genomes = 10).

In this dataset, 58.3% genomes (n genomes = 622; n samples =261) had no AMR determinants detected. There were 176 samples where all *E. coli* genomes had no AMR determinants identified. At least one AMR determinant was found in 66.1% of *E. coli* genomes from chicken, 47.0% of genomes from pork, 31.6% genomes on prawns, 18.2% genomes on salmon and 14.8% of genomes on leafy greens. The AMR determinants were classified into 13 different antimicrobial drug classes, which included eight critically important antimicrobial drug classes and five highly important drug classes (Table 4.2; 4.3; Appendix 4, 5, 6).

Within this dataset, resistance determinants associated with conferring resistance to highest priority critically important antimicrobials included *mcr*-1 (colistin), *bla*CTX-M-27, *bla*CTX-M-55, blaCMY-2, and genes conferring resistance to fluoroquinolones. The AMR determinants *bla*CTX-M-27, *bla*CTX-M-55, blaCMY-2 suggest possible resistance to 3rd-generation cephalosporins and other beta-lactam subclasses.

Table 4.2: The total sample counts and proportions of the collection of 401 samples consisting of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens carrying antimicrobial resistance (AMR) determinants. The AMR determinants are characterised into antimicrobial classes. Clinical importance was determined using the World Health Organization prioritisation categorisation.

Antimicrobial resistance classes	Number of samples with antimicrobial resistance determinants	World Health Organization prioritisation categorisation of clinically important antimicrobials
Polymyxins (Colistin)	2 (0.499%)	Highest Priority Critically Important
Cephalosporins (3rd, 4th-generation)	9 (2.24%)	
Phosphonic acid derivatives (Fosfomycin)	11 (2.74%)	
Beta lactams	141 (35.2%)	Critically important
Aminoglycoside	141 (35.2%)	
Quinolone	83 (20.7%)	
Macrolide	9 (2.24%)	
Oxazolidinone	5 (1.25%)	
Anamycins (Rifamycin)	1 (0.249%)	
Tetracycline	146 (36.4%)	Highly important
Sulphonamide	133 (33.2%)	
Trimethoprim	120 (29.9%)	
Chloramphenicol	46 (11.5%)	
Lincosamide	18 (4.49%)	

Table 4.3: The Distribution of antimicrobial resistance genes across the collection of 401 samples consisting of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens carrying antimicrobial resistance (AMR) determinants.

AMR gene	Number of genomes	Percent of genomes within whole dataset
tet*(A)*	202	18.9
*bla*TEM-1B	187	17.5
sul2	165	15.5
aph(6)-Id	127	11.9
aph(3')-Ib	119	11.2
aadA1	112	10.5
dfrA1	94	8.81
tet(B)	65	6.09
sul1	48	4.50
sul3	37	3.47
dfrA14	35	3.28
floR	31	2.91
qnrS1	30	2.81
cmlA1	27	2.53
aadA2b	25	2.34
dfrA17	22	2.06
dfrA5	22	2.06
*bla*TEM-1C	20	1.87
*bla*TEM-1D	20	1.87
aadA5	19	1.78
lnuF	18	1.69
catA1	17	1.59

Table 4.3: The Distribution of antimicrobial resistance genes across the collection of 401 samples consisting of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens carrying antimicrobial resistance (AMR) determinants.

AMR gene	Number of genomes	Percent of genomes within whole dataset
dfrA12	14	1.31
fosA7	13	1.22
aph(3')-Ia	12	1.12
aph(3')-IId	11	1.03
qnrB19	10	0.937
lnuG	8	0.750
aadA2	7	0.656
*bla*CMY-2	7	0.656
*bla*TEM-1A	7	0.656
tet(M)	7	0.656
*bla*TEM-220	5	0.469
*bla*CARB-2	4	0.375
*bla*OXA-10	4	0.375
mphA	4	0.375
aadA22	3	0.281
dfrA15	3	0.281
dfrA16	3	0.281
mcr-1	3	0.281
qnrS4	3	0.281
ARR02	2	0.187
aadA17	2	0.187
aadA24	2	0.187

Table 4.3: The Distribution of antimicrobial resistance genes across the collection of 401 samples consisting of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens carrying antimicrobial resistance (AMR) determinants.

AMR gene	Number of genomes	Percent of genomes within whole dataset
*bla*TEM-30	2	0.187
catA2	2	0.187
catB3	2	0.187
dfrA36	2	0.187
dfrB4	2	0.187
aac(3)-IV	1	0.0937
aac(3)-VIa	1	0.0937
aadA8b	1	0.0937
aph(4)-Ia	1	0.0937
*bla*CTX-M-27	1	0.0937
*bla*CTX-M-55	1	0.0937
*bla*LAP-2	1	0.0937
*bla*OXA-1	1	0.0937
*bla*TEM-40	1	0.0937
*bla*TEM-99	1	0.0937
dfrA7	1	0.0937
dfrA8	1	0.0937
dfrB1	1	0.0937
erm(42)	1	0.0937
mef(B)	1	0.0937
mph(B)	1	0.0937
qnrB7	1	0.0937

Table 4.3: The Distribution of antimicrobial resistance genes across the collection of 401 samples consisting of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens carrying antimicrobial resistance (AMR) determinants.

AMR gene	Number of genomes	Percent of genomes within whole dataset
qnrS13	1	0.0937
tet(C)	1	0.0937

In this study, 158 *E. coli* positive samples (39.4%) were contaminated with potential MDR *E. coli*. Samples containing at least one MDR genome were found across all food commodities, and were observed in 60.3% of chicken samples, 35.7% of pork samples, 22.7% of prawn samples, 6.06% of salmon samples, and 5.77% of leafy green samples. The proportion of samples containing MDR isolates was significantly different between food commodities (X^2 = 72.5, d.f. = 4, p = 6.66 x 10^{-15}).

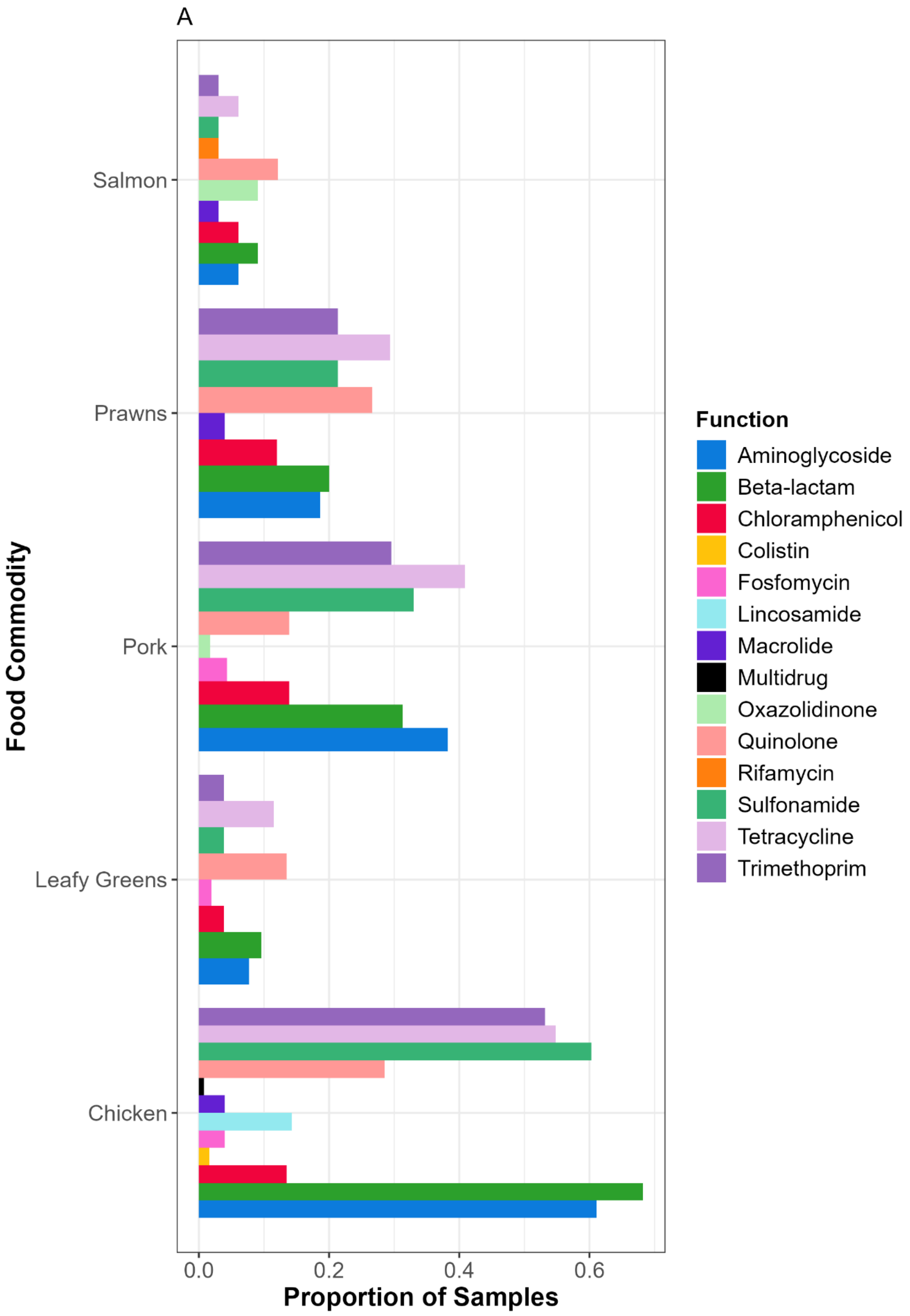
E. coli from chicken meat in this study contained AMR determinants which potentially confer resistance to 12 drug classes and one gene (*erm*(42)) that may confer resistance to multiple classes. *E. coli* isolated from salmon samples contained AMR determinants to 10 drug classes; pork samples had AMR determinants to nine drug classes; leafy greens and prawns had AMR determinants to eight drug classes (Figure 4.5A). Additionally, when up to four isolates per sample were sequenced, up to four different AMR genotypic combinations were often found (Figure 4.5B).

AMR determinants encoding resistance for colistin and 3rd-generation cephalosporins within this dataset were identified across multiple food commodities. Amongst 3rd-generation cephalosporins, *bla*TEM-30 was identified in two putative MDR *E. coli* from two separate chicken samples. Moreover, *bla*TEM-40 was identified in one *E. coli* isolated from a chicken sample where no other AMR determinants were identified. The *bla*CMY-2 AMR determinant was identified in two putative MDR *E. coli* genomes from two separate prawn samples. The *bla*CMY-2 determinant was also found in five *E. coli* genomes from two separate leafy green samples, where no other AMR determinant co-occurred (Table 4.3).

The *bla*CTX-M AMR determinant was identified in two putative MDR *E. coli* genomes from a chicken and prawn sample. The *mcr*-1 AMR determinant was identified in three putative MDR *E. coli* from two chicken samples (Table 4.3).

Table 4.4: The AMR genotypes for *E. coli* genomes carrying AMR determinants encoding resistance for colistin and 3^{rd}-generation cephalosporins from the collection of 401 samples consisting of 1,067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens.

Colistin and 3^{rd} -generation cephalosporins AMR determinant	**AMR genotype**
*bla*TEM-30	aadA1, blaTEM_30, dfrA1, floR, sul1, sul2, tet_A
	aadA1, blaTEM_30, dfrA1, floR, sul1, sul2, tet_A
*bla*TEM-40	blaTEM_40
*bla*CMY-2	aadA1, aadA2b, blaCMY_2, blaTEM_1B, cmlA1, dfrA12, mef_B, sul3, tet_B
	aadA1, aadA2b, blaCMY_2, blaTEM_1B, cmlA1, dfrA12, mef_B, sul3, tet_B
	blaCMY_2
*bla*CTX-M	blaCTX_M_55, blaTEM_1B, mph_A, qnrS13, tet_A
	aadA1, aph_3____Ib, aph_6__Id, blaCTX_M_27, dfrA1, floR, sul1, sul2, tet_A, tet_B,
mcr-1	aph_3____Ib, aph_6__Id, blaTEM_1B, mcr_1, sul2
	aph_3____Ib, aph_6__Id, blaTEM_1B, mcr_1, sul2,
	aph_3____Ib, aph_6__Id, blaTEM_1B, mcr_1, sul2,


A
Food Commodity
Salmon
Prawns
Pork
Leafy Greens
Chicken
0.0
0.2
0.4
0.6
Proportion of Samples
Function
Aminoglycoside
Beta-lactam
Chloramphenicol
Colistin
Fosfomycin
Lincosamide
Macrolide
Multidrug
Oxazolidinone
Quinolone
Rifamycin
Sulfonamide
Tetracycline
Trimethoprim

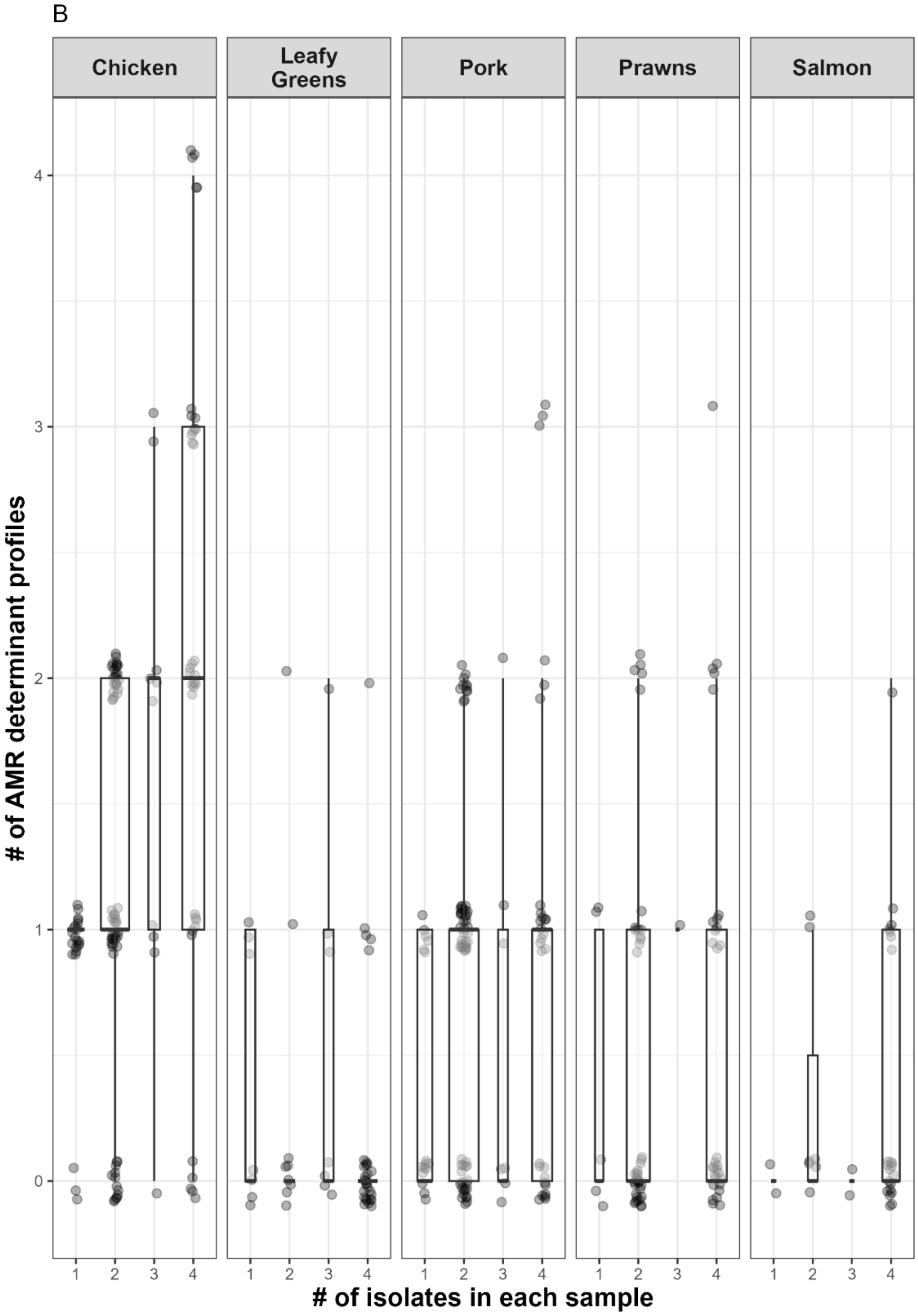


Figure 4.5: The antimicrobial resistance determinants conferring potential antimicrobial resistance (AMR) found within the collection of 1,067 *E. coli* genomes isolated from a total 401 retail chicken, pork, prawns, salmon, and leafy green samples. A) The proportion of samples for each food commodity containing at least one AMR determinant within each food commodity. AMR determinants classed into antimicrobial resistance drug class. B) Boxplots of the 1,067 *E. coli*, showing the number of unique combinations of AMR determinants (termed AMR profiles) recovered in all 401 food samples when up to four *E. coli* genomes were sequenced, separated by food commodity. Each dot is one sample.

4.4.6 Phylogeny of *E. coli* contaminating retail foods

The *E. coli* phylogenetic tree was divided into phylogroups A, B1, B2, C, D, E, F and G. *E. coli* genomes classified into phylogroups B1 (40.9%) and A (34.0%) comprised the majority of the collection. Phylogroups A, B1, B2, and D contained *E. coli* from all food commodities. Between 0-14 unique AMR genes and point mutations were identified within individual genomes across the whole dataset. The 26 presumptive ExPEC were identified across six phylogroups (Figure 4.6).

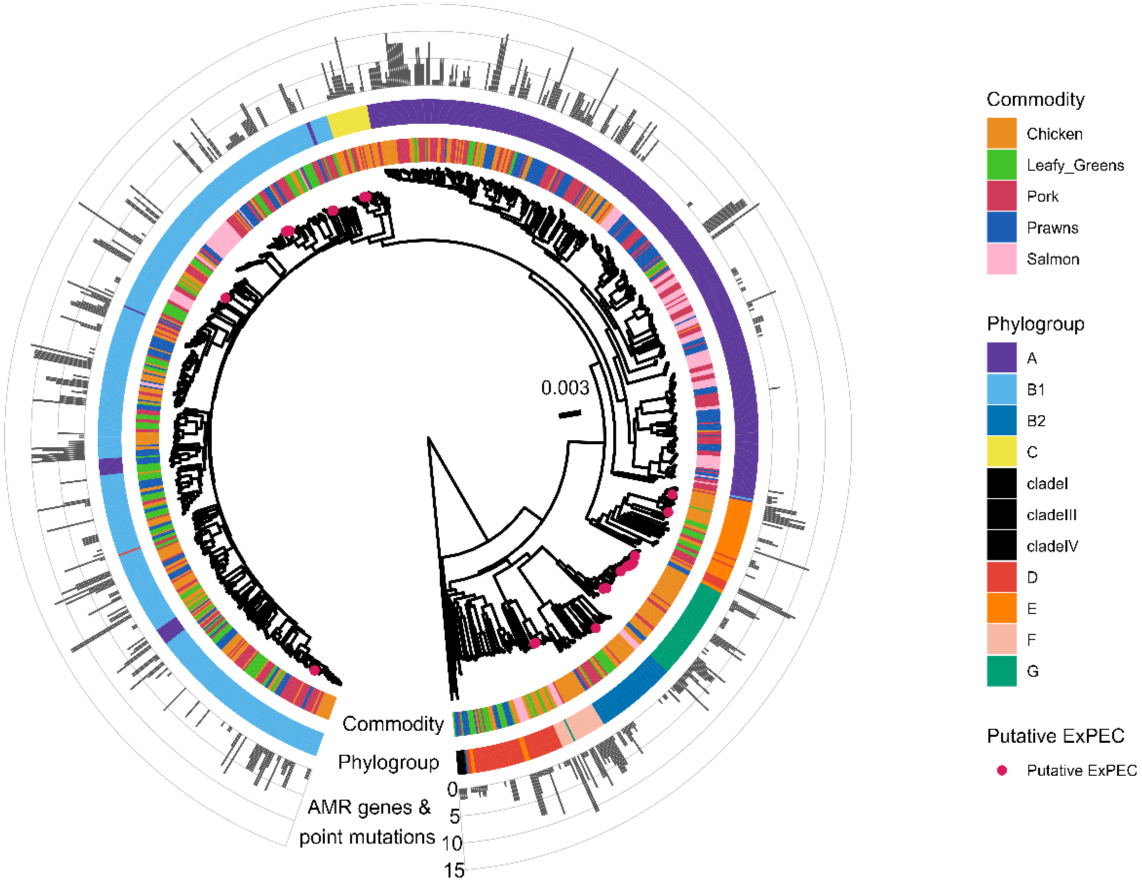


Figure 4.6: Phylonium midpoint rooted phylogenetic tree the collection of 1067 *E. coli* genomes isolated from retail chicken, pork, prawns, salmon, and leafy greens with a tree scale estimating 0.003 nucleotide substitutions per site. Tree tips are labelled red if they are putative ExPEC. The commodity (inner ring), phylogroup (middle ring), and the total number of AMR genes and point mutations are mapped onto the tree (outer ring).

Two clades were examined due to the enrichment of isolates originating from chicken and salmon commodities on the phylogeny. Within the phylogeny, 53 genomes from 39 samples were classified as phylogroup G. The genomes in phylogroup G were classified into 10 unique STs. The majority of phylogroup G was comprised of *E. coli* isolated from chicken (n genomes from chicken = 44/53; n chicken samples = 34/39) and majority of genomes were classified as ST117 (n genomes classified as ST117 = 41/53). Majority of the food samples ST117 contaminated was chicken meat (n ST117 genomes = 37/41). The number of AMR determinants in each genome from phylogroup G ranged from 0-8 and 28 samples had *E. coli* with at least one AMR determinant (Figure 4.7). A different subclade within phylogroup B1 comprised entirely of *E. coli* isolated from salmon (n genomes = 19; n samples = 7). All the genomes in this clade were all classified as ST5474. These genomes had no AMR determinants detected.

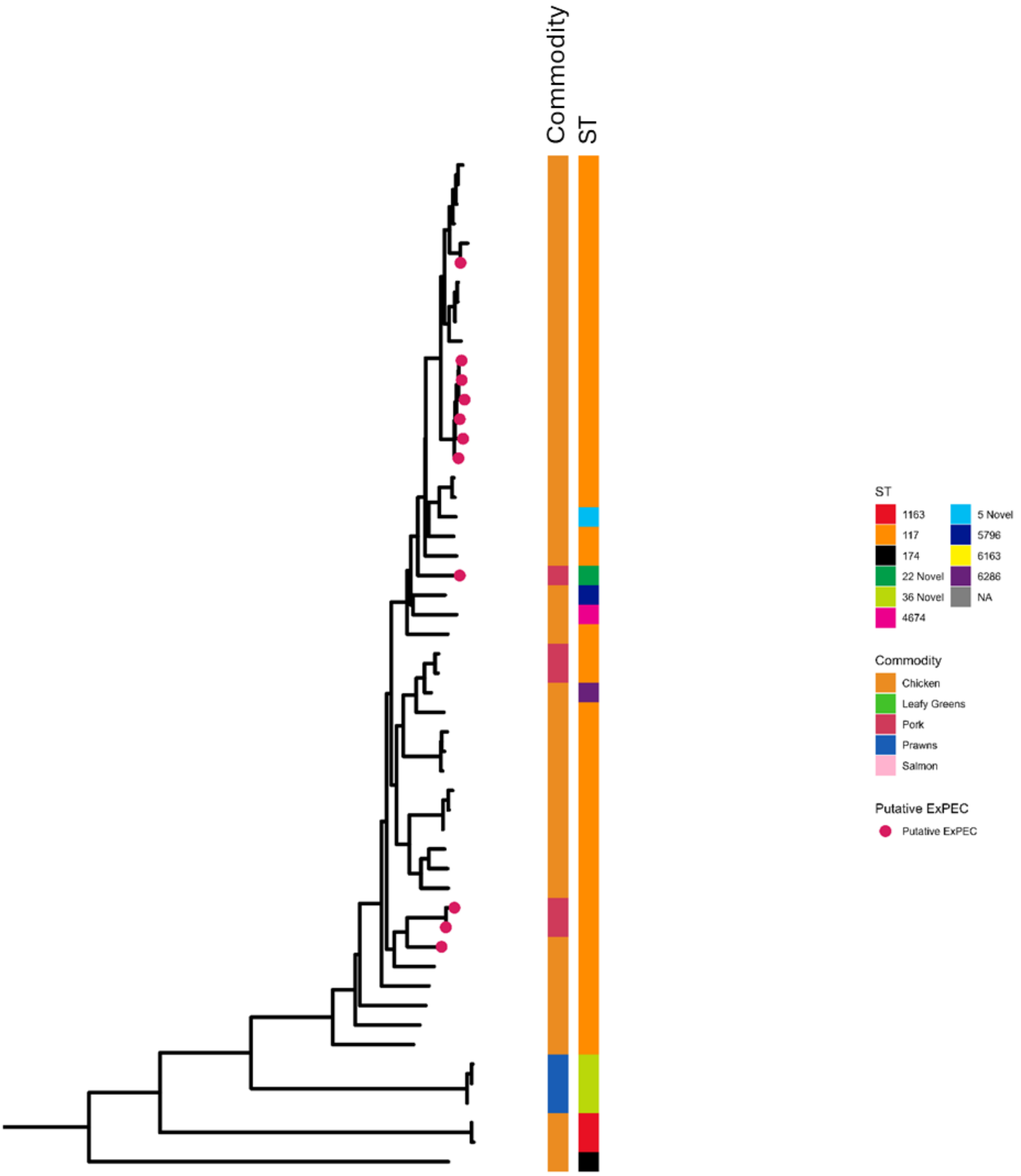


Figure 4.7: A clade of the Phylonium midpoint rooted phylogenetic tree comprising of primarily *E. coli* genomes from chicken meat samples from phylogroup G, highlighting a subset of genomes from the collection of 1,067 *E. coli* isolated from retail chicken, pork, prawns, salmon, and leafy greens. The tree scale is estimated to be 0.003 nucleotide substitutions per site. Tree tips are labelled red if they are putative extra-intestinal pathogenic *E. coli* (ExPEC). The commodity (left column) and sequence type (right column) are labelled.

4.5 Discussion:

WGS of multiple isolates per sample revealed that heterogenous populations of *E. coli* contaminate retail food samples from different food commodities. These results indicate that the sampling up to four *E. coli* even within a sample can have different STs across the phylogeny, and different accessory genes. This has implications for food safety as *E. coli* enumeration are one of the standard approaches for identifying unsafe levels of faecal contamination in the UK (International Organization for Standardization, 2015). When *E. coli* enumeration is used alone, it does not show the different contaminating types of *E. coli*. By investigating the diversity of *E. coli* on food and within individual food samples, more information is available for assessing the risk of current and new potential threats to the consumer.

4.5.1 There are limited WGS *E. coli* ST diversity studies focusing on the *E. coli* population within a sample

In this study, the 1,067 *E. coli* isolates were classed into 238 different known and 54 unknown STs. Similar studies in the United States investigating raw chicken, turkey and pork classified 117 *E. coli* into 81 known STs (Aworh *et al.*, 2024), and another retail meat survey in China assigned 330 *E. coli* isolates into 146 different known STs and 20 novel STs (Wang *et al.*, 2024). Generally, our study is in line with previous work, which show retail meats are contaminated by large number of unique STs. Despite similarities with how meat samples were collected and how *E. coli* was cultured from meat samples between this study and these two previous studies, the work in this chapter took up to four isolates, whereas the other studies only took one isolate forward for whole genome sequencing (Aworh *et al.*, 2024, Wang *et al.*, 2024). This means that the within-sample ST diversity could not be explored within these previous retail meat studies and that the number of STs found within this study are not necessarily comparable, as there is a higher chance of missing out a rarer ST. The study in China also used the core genome MLST (cgMLST) approach, which uses hundreds of gene loci and a predefined allele scheme to assign numbers for each locus. Due to the differences between schemes used, there cannot be direct comparison between cgMLST and traditional MLST typing schemes since novel allele numbers are not synced.

Regarding leafy greens, this work found a higher ST richness across the study population (68 STs in 162 isolates) compared to a German study evaluating tetracycline-resistant *E. coli* from leafy vegetables (28 STs in 120 isolates) (Reid *et al.*, 2020). Additionally, the three most abundant STs by Reid *et al.*, were not found to contaminate any leafy greens in this study. The difference in number of STs found may be because Reid et al., were specifically investigating tetracycline-resistant *E. coli*, whereas our work did not. Moreover, the disagreement in STs between studies may also be due to the differences in farm locations, processing procedures and the potential contamination sources associated with both (i.e. soils, manure usage, farmers/factory workers).

A limited number of *E. coli* whole-genome surveys have been done on seafoods. A tetracycline-resistance survey on a range of seafoods from China found that nine *E. coli* isolates were classed into eight known STs and one unclassified ST (Liu *et al.*, 2024). Another survey on retail fish in Tunisia found that six isolates were classed as four different STs (Sola *et al.*, 2022). There are a lack of studies looking specifically at *E. coli* contamination on salmon and prawns. Therefore, comparison of these general seafood studies to this dataset is limited, especially as there are major differences between the geographical regions, types of markets, seafood, and total number of *E. coli* sequenced. ST58 and ST10 were found in both this study and the Chinese study but have both been described as generalists (McKinnon *et al.*, 2018, Manges and Johnson, 2012). The lack of general surveillance on leafy greens and seafood highlights this research as key to setting a baseline for future risk assessment and surveillance studies across these food commodities. However, further WGS surveillance is required to investigate whether there are patterns of sequence type contamination across different retail food commodities.

4.5.2 Taking multiple isolates per sample reveal different sequence types

When up to four isolates were taken per sample, these isolates could be classified into four different STs. A similar study in the US, subtyped five *E. coli* colonies per sample using Enterobacterial Repetitive Intergenic Consensus Polymerase Chain Reaction (ERIC2-PCR) and found that up to five distinct genotypes could contaminate chicken and pork (Yamaji *et al.*, 2018). ERIC2 does not provide the same resolution as whole genome sequencing, but is still a useful approach for investigating the different types of *E. coli* that contaminate food (Alsultan and Elhadi, 2022). Another study previously used repetitive element palindromic (REP) PCR screening found that up to nine *E. coli* REP-types could be recovered from a single meat sample (Vangchhia *et al.*, 2018).

In this study, salmon had a maximum of two STs recovered per sample. Salmon was previously found to have a lower prevalence of *E. coli* than the other food commodities (Janecko *et al.*, 2023), suggesting that there may be factors affecting the ability of *E. coli* to persist on salmon. This could include a lower likelihood of recovery of *E. coli* from salmon using the method described here, or fewer opportunities for *E. coli* contamination during the salmon processing chain. Salmon muscle has been suggested to be sterile, but exposure to the water environment (Austin, 2006) and the food manufacturing process (Møretrø *et al.*, 2016) can lead to bacterial contamination. These studies and our work suggest that taking four isolates is insufficient for capturing the full range of *E. coli* STs contaminating retail foods, with salmon as the exception.

4.5.3 There is within-sample within-ST SNP variation

In this study, several samples had *E. coli* classified as the same ST on the same sample with a pairwise non-recombinant SNP difference greater than five SNPs. Previous work investigating ST131 (Forde *et al.*, 2019), and ST117 (Xia *et al.*, 2022) indicate that *E. coli* from a single ST are not all clonal, and a theoretical ancestor can accumulate SNPs and lead to different lineages within the same ST. Within-ST SNP variation in *E. coli* has been observed in human and environmental studies previously (Sonda *et al.*, 2018, Vingino *et al.*, 2021, Hazen *et al.*, 2023) and variation within STs found within the same samples has also been observed in non-typhoidal *Salmonella*, *Yersinia enterocolitica,* and *Pseudomonas aeruginosa* from retail foods in the UK (Bloomfield *et al.*, 2023a, Palau *et al.*, 2024, Bloomfield *et al.*, 2024). As within-sample within-ST SNP variation is evident for the top three most abundant STs, it is likely that there are other STs in this collection that will differ by greater than five SNPs.

The food commodities in this study have different growth and rearing timelines. According to the Department of Environment, Food and Rural Affairs, the age of slaughter for indoor-system chickens is roughly 35 days (Royal Society for the

Prevention of Cruelty to Animals, 2022); extensive-indoor or barn-reared chickens is a minimum of 56 days or a minimum of 81 days for free-range chickens (Department for Environment Food and Rural Affairs, 2024). Pigs take longer to rear than chickens and are slaughtered when they reach market body weight, which is around 175 days (López-Vergé *et al.*, 2018). According to the Commonwealth Scientific and Industrial Research Organisation (CSIRO), farmed prawns are harvested around five to seven months of growth (Commonwealth Scientific and Industrial Research Organisation, 2019), while leafy greens (Jasper *et al.*, 2021) are estimated to have individual growing timelines. Food manufacturing processes vary between commodities and also depend on the cuts of meat/end product, but for all commodities, these individual processes take less than a day (GOV UK, 2009, British Meat Processors Association, 2022, Leschen, 2005, Jasper *et al.*, 2021). From packaging to distribution, the Food Standards Agency's Advisory Committee on the Microbiological Safety of Food state that vacuum packed foods should have a shelf life of ten days (Advisory Committee on the Microbiological Safety of Food, 2020). *E. coli* ST131 was estimated to accumulate around 6.50 x 10^{-7} substitutions per site per year, which equates to 2.7 SNPs per year per genome (White *et al.*, 2024) The SNP differences observed within this dataset suggests they could not accumulate this difference in this time.

The presence of diverse *E. coli* may occur from multiple cross-contamination points throughout the food processing chain, for example conveyor belts and weighing units for leafy greens (Holvoet *et al.*, 2012), or machinery and equipment in meat processing facilities (Zwirzitz *et al.*, 2020). Additionally, a single cross-contamination event from the food manufacturing process could introduce multiple lineages of *E. coli*. For example, a study in Japan identified different STs contaminating the same location in a food factory (Nakamura *et al.*, 2021). Microbial contamination can occur at any point along the food chain, including at production, processing or during food retail and distribution; as the sampling for this study was only performed at retail level, there is not enough information to determine when the retail foods were contaminated and by which source (Rouger *et al.*, 2017, Ludden *et al.*, 2019). Initial contamination may have also occurred from the original animal host, contaminating the food processing chain and leading to the variety of STs observed contaminating the end product. This has been observed for *E. coli* O157, where *E. coli* from the slaughtered cattle survived the food processing chain and led to an outbreak (Wilson *et al.*, 2018). Leafy vegetables have an alternative initial contamination pathway, where soil and manure could be the source of the contamination (Islam *et al.*, 2004).

However, this approach comparing genomes to a reference and comparing SNP differences has limitations. This is a stringent reference-based approach that will likely underestimate the number of SNPs between genomes. As genomes classed as the same ST are not necessarily clonal and can be comprised of different lineages that have diverged from a theoretical ancestor, using one reference is likely to introduce some bias due to the genetic variability within STs (Decano and Downing, 2019). However, it is still a useful estimation for observing low and high SNP differences in *E. coli* and demonstrates an even higher level of potential contamination and risk.

4.5.4 No diarrheagenic *E. coli* were found, but presumptive ExPEC were identified

Within this dataset, there is likely a low risk of consumers being exposed to diarrheagenic *E. coli* as no *stx1, stx2*, *eae, bfp, let, estA/estB,* nor the Afa/Dr family virulence genes were present. However, defining diarrheagenic *E. coli* based on virulence genes is challenging as clinical presentation of these infections can vary, which is evident by asymptomatic hosts that can harbour diarrheagenic *E. coli* with virulence genes associated with pathogenicity (Jesser *et al.*, 2023). Further complications for defining diarrheagenic *E. coli* arise because of the presence of pathotype hybrids (Kislichkina *et al.*, 2022). An example of which had both the aggregative adhesive phenotype and virulence genes indicative of enteroaggregative *E. coli*, as well as the genes indicative of Shiga-toxin-producing *E. coli* (Bielaszewska *et al.*, 2011).

A set virulence gene scheme by Johnson *et al.*, (Johnson *et al.*, 2003) defined 26 genomes in this dataset as presumptive ExPEC. Contaminated retail foods have been suggested to be a reservoir for ExPEC (Yamaji *et al.*, 2018), but it is unlikely to be a frequent direct transmission route (Day *et al.*, 2019). Additionally, the absence of detectable putative ExPEC on leafy greens, salmon and prawns does not exclude them as potential vehicles for pathogenic lineages; leafy greens can be a vehicle for pathogenic *E. coli* (Mikhail *et al.*, 2018), and may present a potentially greater risk than the other food commodities as leafy greens are typically eaten raw.

4.5.5 *E. coli* contaminating food can have different AMR profiles

This study identified AMR determinants that could potentially confer antimicrobial resistance to seven critically important clinical antimicrobials (World Health Organization, 2019, Catalano *et al.*, 2022), revealing *E. coli* positive samples with beta-lactam resistance genes (36.2%) and aminoglycoside resistance genes (35.2%) present. Extended spectrum beta-lactamase producing *E. coli* have been commonly reported in retail foods and have been suggested that *E. coli* in general are reservoirs for ESBL dissemination (Day *et al.*, 2019). Colistin resistance genes (*mcr-1*) were also found in two genomes, which is concerning as this is a last-resort treatment for multi-drug resistant Gram-negative pathogens that can potentially be shared with other bacteria (Mohapatra *et al.*, 2021). This risk is also elevated by potential MDR *E. coli*, which were identified in 33.4% of samples within this dataset. AMR genes can be mobile and have been shown to spread between different *E. coli* phylogroups, and across species to *Salmonella* Typhimurium *in vitro* and within *in vivo* mice models (Benz *et al.*, 2021). The potential spread of these genes to other bacteria and other colonising *E. coli* on food means the threat is not limited to just which *E. coli* can successfully colonise and cause disease, but also the *E. coli* that harbour mobile genes of interest.

Previous work investigating microbial communities estimated that retail foods can be contaminated with varying concentrations of AMR (Bloomfield *et al.*, 2023b). This is important as *in vitro* gut model studies have identified that consumption of high bacterial loads with AMR genes can survive the digestive process and multiply (Buberg *et al.*, 2023), as well as spread AMR determinants (Lambrecht *et al.*, 2019). Although the bacterial load on retail foods will likely be less than the *in vitro* studies overall, and certain foods like raw meats and seafood will be cooked which will reduce overall bacterial load and AMR exposure. This risk is also food commodity dependent, as leafy greens may not be cooked before eating and will therefore pose a different overall health risk.

Multiple AMR profiles within a sample highlight the need to take multiple isolates as one isolate from each sample underestimates the risk to the consumer. Potential multidrug resistant *E. coli* was identified in 33.4% of the *E. coli* positive samples collected, of which chicken samples had the highest proportion of samples containing potentially MDR *E. coli* across the five commodities. Raw chicken meat are a potential reservoir for MDR *E. coli* and according to the FSA, poultry, specifically chicken and turkey, were the first priority for AMR surveillance in 2023 (Nicholls *et al.*, 2023). Our study highlights that MDR *E. coli* can contaminate all these food commodities, which may have been excluded if there was a selection criterion for a specific resistance to an antibiotic drug class. It should be noted that the presence of these genes do not mean they will be phenotypically expressed – for example the presence of *tet(A)* and *tet(B)* genes had poor agreement with actual display of tetracycline resistance from *E. coli* isolates in clinical settings (Hussain *et al.*, 2014). Although phenotypes are required to confirm expression of AMR and MDR, genotypic data remains a useful indicator of AMR genes exposed to the consumer, which can help inform public health officials with proactively reducing AMR on food.

The co-occurrence of the IncFIB plasmid replicons with AMR determinants warrants further investigation into whether plasmids carrying AMR determinants are present within the *E. coli* in this dataset. The predominance of IncFIB replicons within *E. coli* isolated from chicken samples aligns with previous work which identified IncFIB as the predominant plasmid replicon in *E. coli* from poultry and retail poultry meat (Johnson Timothy *et al.*, 2007). IncF plasmids are the most abundant amongst *E. coli* and are known to be linked with β-lactamases, *tet* and *sul* genes. As there is a metabolic cost associated with plasmid maintenance, the predominance in the poultry commodity suggests that there may be sufficient positive selection pressures to maintain the plasmids. This highlights retail chicken meats as an important route of *E. coli* dissemination mobile genetic elements with IncFIB and AMR determinants (Pitout and Chen, 2023).

4.5.6 *E. coli* genomes from different commodities are intermixed across the phylogeny, with limited exceptions

Broadly across the phylogeny, the *E. coli* genomes from different food commodities are intermixed, with some exceptions. The ST117 and ST5474 clades were chosen due to the enrichment of *E. coli* from the chicken and salmon food commodities within the phylogeny. The enrichment of these specific *E. coli* was investigated to determine whether or not there was a link with population structure. Phylogroup G was identified to have *E. coli* genomes from primarily chicken meat cluster together. The majority of *E. coli* were classified as ST117, which is a globally distributed potential ExPEC lineage associated with poultry and has been highlighted to consistently contaminate chicken meats (Xia *et al.*, 2022, Ronco *et al.*, 2017). The variability of AMR determinants within these ST117 genomes in this study and their potential as putative ExPEC highlights these *E. coli* genomes as clones of concern, especially when exposed to a consumer.

A salmon specific clade in phylogroup B1 was also identified on the phylogeny but did not contain any AMR determinants. The presence of *E. coli* on farmed fish is suggested to be influenced by the environmental factors like water temperature and pH (El Badawy et al., 2025) and for wild caught fish, the fishing vessel can be another point of contamination introducing different lineages of *E. coli* (Rosen and Levin, 1970). This clade was comprised of ST5474 and limited research has identified this ST on hospital flies (Heiden et al., 2020) and in diarrheal stool samples (Begum Yasmin et al., 2018). There may be clinical implications of this ST, but it is unknown if this lineage will pose a health threat to consumers.

Whilst the *E. coli* genomes were broadly assigned to expected phylogroups, the phylogeny showed some interspersion of genomes from different phylogroups within neighbouring clades. This interspersion may reflect methodological limitations rather than true phylogenetic signal. The phylogroup typing software relies on four genes and is sensitive to assembly fragmentation which can therefore affect the phylogroup assignment (Beghain *et al.*, 2018). This may result in *E. coli* genomes being classified as an inconsistent phylogroup within a clade within the phylogeny. Alternatively, this may also be due to interference by mobile genetic elements, as horizontal transfer events have previously been argued to affect phylogroup assignment (Beghain *et al.*, 2018).

4.5.7 The culturing process may limit growth of all types of *E. coli*

This study focused on culturing the general population of *E. coli* without a strict focus on pathogenic and antimicrobial resistant lineages. The culturing process used in this study is standard for culturing *E. coli* on meats, but the initial 42-degree incubation temperature could potentially select for certain *E. coli* lineages that can tolerate this temperature. For example, similar methodologies using the ISO 16649-1:2018(en) method to enumerate beta-glucuronidase-positive *E. coli* actively selects against pathogenic *E. coli* O157:H^- types and other pathogenic lineages at 44 degrees (International Organization for Standardization, 2015). Naturalised *E. coli* from the environment have been grown at lower 30–37-degree temperatures in previous surveys (Devane, 2019, Moinet *et al.*, 2021, Ishii *et al.*, 2006), but have not determined the heat tolerances of these lineages. Regardless, previous work indicates this approach is useful for investigating faecal coliforms, which grow well at 44.5 degrees (Korajkic and Harwood, 2016), meaning *E. coli* from non-environmental host sources contaminating these foods are likely to thrive from this methodology.

4.6 Conclusions

The variation in virulence and AMR genes in combination with the range of diverse lineages on the phylogenetic tree suggests that consumers are potentially exposed to a variety of *E. coli* lineages with different levels of risk. Despite defined satisfactory CFU and MPN levels on food (UK Health Security Agency, 2023a), the number of cells required to cause infection is relative to the lineage of *E. coli* (Kothary and Babu, 2001). To develop the current safety policies, WGS multi-isolate approach for surveillance is needed to monitor public health risks. Further work needs to assess contamination at the retail level across an international scale and across multiple time points as food importation and globalisation occur. A One Health approach that considers animals, the food processing environment and humans that interacts within these environments, is essential for understanding the diversity of *E. coli* contamination.

To our knowledge, this is the most comprehensive multi-isolate WGS study of *E. coli* contamination on retail foods. This chapter found that there are a diverse population of *E. coli* contaminating samples that are not revealed with current enumeration methods. This provides an incomplete assessment of risk. WGS multi-isolate approaches are therefore necessary to identify AMR and health risks contaminating retail foods.

As this chapter indicates, the contamination observed on food is comprised of a heterogeneous population of *E. coli* which is likely an indicator of multiple contamination points or the introduction of multiple *E. coli* lineages at one point. Due to the nature of this dataset, it is not possible to definitively say the original host of the *E. coli*, when they were contaminated and how they have adapted to survive and thrive on retail foods. As this environment presents with too many external factors which may influence the population, the next chapter will focus on *E. coli* directly isolated from hosts.

Chapter 5: Evaluating inter-species *E. coli* co-occurrence within humans and companion dogs cohabiting in dog-owning households

Chapter contributions: Design of original survey and sample collection was done by Menezes, J., Pomba, C., Frosini, S.M., and Loeffler, A., in collaboration between the University of Lisbon and the Royal Veterinary College; Study design, growth, detection and isolation of the *E. coli* were done by Astorga, G and Frosini, S.M.; Scripts, code, and bioinformatic analysis were written by Astorga, G and were discussed with Tiwari, S.K., and Savva, G., Mather, A.E., Bloomfield, S. J., and de Oliveira Martins, L.; The genome wide association study was discussed by Astorga, G., Tiwari, S.K., and Savva, G.; Astorga, G., Mather, A.E., Frosini, S.M., Singh, D., Wain, J., Bloomfield, S.J., Savva, G., Tiwari, S.K., and de Oliveira Martins, L. all discussed the work.

Publication: This chapter is being prepared for publication under the title “Whole genome comparative analysis shows limited *Escherichia coli* co-occurrence within co-habiting humans and dogs” for BMC Genomics.

5.1 Introduction:

E. coli is a common gut coloniser and the human gut has been suggested to harbour a mixture of different types of residential and transient *E. coli* (Foster-Nyarko and Pallen, 2022, Sears *et al.*, 1950). One route of exposure may be from the environment, such as food as discussed in chapter 4. However, humans can also be exposed to *E. coli* from sources like companion animals, which may harbour *E. coli* and can then be shared through close contact (Muloi *et al.*, 2022). This may also be true for the reverse, as there is potential for human-associated *E. coli* lineages to appear in companion animals (Pomba *et al.*, 2009, Ewers *et al.*, 2010).

The most popular companion animal in the UK was the domesticated dog (Anderson *et al.*, 2023), *Canis lupus familiaris*. In the UK there is estimated to be 13.5 million dogs kept as pets, with 36% of households having a dog (UK Pet Food, 2024). The proximity within households and repeated contact may promote *E. coli* sharing and long-term colonisation as indicated by macro-restriction PFGE analysis (Johnson *et al.*, 2008, Johnson and Clabots, 2006).

Humans and companion dogs can be asymptomatic carriers of similar *E. coli* as determined by identical macro-restriction PFGE results, AMR phenotypes and genotypes (Naziri *et al.*, 2016, Stenske *et al.*, 2009). Higher resolution WGS studies have further identified similar *E. coli* present in humans and companion dogs with AMR genes of concern like extended-spectrum beta-lactamase genes (Menezes *et al.*, 2023, Menezes *et al.*, 2022, van den Bunt *et al.*, 2020, Cozma *et al.*, 2022). MLST and phylogeographic investigative approaches have also inferred transmission between human and canine hosts (Naziri *et al.*, 2022, Johnson *et al.*, 2022).

Identification of bacterial transmission (indicating direction) and the lesser used term sharing (unknown directionality) between hosts requires longitudinal data and thorough epidemiological data. For cross-sectional studies, potential transmission or sharing cannot be definitively determined, but previous work using macro-restriction PFGE identified co-occurrence between humans and dogs (Harada *et al.*, 2012, Stenske *et al.*, 2009, Naziri *et al.*, 2016).

Lower resolution macro-restriction PFGE approaches have identified genetically similar *E. coli* between humans and dogs within households, but these methods are not as discriminatory as WGS. Whilst studies have utilised WGS to investigate *E. coli* co-carriage in humans and dogs, there has been a focus on selecting for AMR phenotypes (Menezes *et al.*, 2023, van den Bunt *et al.*, 2020, Cozma *et al.*, 2022). This selective approach limits the observations on *E. coli* diversity and possible co-occurrence between hosts. Additionally, WGS can help assess the risk of both humans and companion animals as a source of *E. coli* with AMR.

5.2 Aims and objectives

The objectives for this chapter were to:

i) Investigate if putative clonal *E. coli* co-occur within different hosts in this dataset.

ii) Use comparative genomics to identify similarities and differences within *E. coli* isolated from different hosts.

iii) Identify AMR determinants and extra-intestinal pathogenic status indicating virulence factors within the dataset.

iv) Assess the risk to humans and dogs from potential sharing of *E. coli*

5.3 Materials and methods

5.3.1 Human and companion dog study population

Faecal samples were collected following signed, informed consent prior to enrolment as part of a previous longitudinal study at the Royal Veterinary College (RVC), Hertfordshire, UK in collaboration with the Centre for Interdisciplinary Research in Animal Health, Faculty of Veterinary Medicine, University of Lisbon, Lisbon, Portugal (Menezes *et al.*, 2023). Ethical approval was obtained from the RVC's Clinical Research Ethical Review Board (CRERB; URN 2017 1750-3).

This previous longitudinal study included UK households with at least one human and at least one companion dog living in close contact using convenience sampling. All human participants consented to the main goals of the study. An epidemiological questionnaire was completed by human members of each household (Menezes *et al.*, 2023). Companion animals' health was evaluated by veterinary professionals supplemented by behavioural data provided by the owner (Menezes *et al.*, 2023).

5.3.2 Sample collection and processing

Human and animal faecal samples were collected by owners using instructions provided by the RVC. Human faeces were collected in sterile 'FeCol' collection papers (Alpha Laboratories Ltd, UK), transferred into sterile plastic containers, or directly into sterile plastic containers. Partial dog faeces that did not touch the ground were transferred into sterile plastic containers. An aliquot of each faecal sample was stored in 0.9% NaCl (Merck, Germany) and 50% glycerol (Sigma-Aldrich, USA) in the RVC biobank at -70/-80 °C.

Prior to growth, faecal samples were defrosted and homogenised. Samples were enriched in a 1:1 ratio of buffered peptone water (Biokar Diagnostics, France) and incubated at 37 °C for 18±4 h. All subsequent incubations were performed at 37 °C for 18±4 h. A 1:10 serial dilution was performed using 0.9% saline and 100µl of each dilution was plated on Tryptone Bile X-Glucuronide (TBX) agar (Oxoid, Basingstoke, UK) and incubated.

Up to eight colonies with different morphologies from each TBX plate were taken. The typical *E. coli* morphology is a clear circular convex blue/green coloured colony. Morphologies that consistently exhibited other characteristics: a less vibrant, but still blue/green sheen; or a more mucoid colony with a blue/green sheen; or a non-circular colony with a blue/green sheen were also chosen.

The presumptive *E. coli* colonies on a plate were streaked onto 5% sheep's blood agar (Thermo Fisher Scientific, Basingstoke, UK) and incubated. Colonies were biochemically confirmed as presumptive *E. coli* using a negative Simmon's citrate agar (Sigma-Aldrich, Haverhill, UK) and positive Kovacs indole spot reagent test (Fisher Scientific, Loughborough, UK). Biochemically confirmed colonies were streaked onto Mueller-Hinton agar and incubated, stored on Amies transport swabs (Scientific Laboratory Supplies, Nottingham, UK) and transported to Quadram Institute Bioscience. There, each colony was subcultured onto tryptic soy agar (Trafalgar Scientific Ltd., Leicester, UK) and were stored at −70 °C in 1 mL of Brucella broth +17.5% glycerol (Thermo Fisher Diagnostics, Rochford, UK).

A household would be termed "complete" and used for this study according to the following inclusion criteria:

a) There was at least one healthy human and at least one healthy dog cohabiting.
b) Faecal samples were available from the same timepoint from both humans and dogs within the same household.
c) There was at least one *E. coli* isolated from one healthy human and one healthy dog in a household.

Only *E. coli* from complete households were used for downstream analysis. Metadata for all *E. coli*-positive hosts from complete households were obtained and used for this study.

5.3.3 Whole genome sequencing, quality control and assembly

Illumina short read whole genome sequencing and library preparation were previously described in **Section 2.3**. Post whole genome sequencing, isolated *E. coli* colonies are referred to as genomes. Genomic analyses were performed on the High-Performance Computing (HPC) cluster at the Norwich Biosciences

Institutes on a Quadram Institute Bioscience QIB-CLOUD Virtual Machine. Quality control metrics are described in chapter 2.

5.3.4 *E. coli* phylogroup, sequence typing, and phylogenetic analysis

E. coli genome assemblies were classified into phylogroups and sequence types (STs) as stated in **Section 2.4.2.** Unclassified STs were further characterised into matching novel STs by identifying unique combinations of *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA* profiles provided by the MLST output (Appendix 3). Serogroups were determined using ECtyper 2.0.0 using default parameters on assemblies (Bessonov *et al.*, 2021).

As outlined in chapter 3, Phylonium v.1.6 (Klotzl and Haubold, 2020) with default settings was used to estimate the phylogenetic distances between the *E. coli* assemblies. A dendrogram was estimated using rapidNJ v.2.3.2 (Simonsen *et al.*, 2008) and was visualised using R v.4.1.3 (R Core Team, 2022).

5.3.5 Antimicrobial resistance determinants and virulence genes

The presence of AMR genes, plasmid replicons and virulence genes were determined using ARIBA v.2.14.6 (Hunt *et al.*, 2017) with default settings on the trimmed reads. The Resfinder (Bortolaia *et al.*, 2020), PlasmidFinder (Carattoli and Hasman, 2020), and VFDB (virulence factor database) core (Liu *et al.*, 2021) databases were downloaded in May 2023. Known point mutations associated with AMR were identified using PointFinder (python script downloaded in October 2024) (https://github.com/guthrielab/pointfinder) (Zankari *et al.*, 2017).

AMR genes and point mutations are collectively termed AMR determinants. The AMR determinants were categorised into their antimicrobial classes using the Comprehensive Antibiotic Resistance Database (CARD) (McArthur *et al.*, 2013). The AMR determinants were classified by differing priority levels according to the World Health Organization scheme (World Health Organization, 2024). The criteria to classify the *E. coli* as potential multidrug-resistance (MDR) in this study is stated in **Section 2.4.3**.

E. coli genomes classified as the same ST with the same AMR determinant combination (termed AMR genotypes) from each host were termed “replicates”. To identify total counts of AMR determinants, replicates were removed. A Wilcoxon

Rank Sum Test was used to investigate if the number of unique AMR determinants in *E. coli* differed between host species.

E. coli isolates were characterised as extraintestinal pathogenic (ExPEC) according to the requirements stated in **Section 2.4.3**.

5.3.6 Pairwise SNP analysis

E. coli STs of interest for pairwise SNP analysis were selected for a reference-based approach via two ways:

a) Genomes classified as the same ST that were present in both humans and dogs in this dataset were selected.

b) A core gene alignment was created using Roary v. 3.12.0 (Page *et al.*, 2015), where core genes were defined as genes present in 95% of genomes with the minimum percent identity to cluster homologous genes set to 90%. The pairwise SNP differences were calculated using SNPdist v.0.7 (https://github.com/tseemann/snp-dists). Genomes that were identified as: a) The same ST; b) less than five SNPs different using this core gene alignment SNP approach; and c) present in different hosts were selected.

E. coli genomes from the highlighted STs by these two approaches were grouped into their respective ST classifications. A reference genome for each ST was chosen using Referenceseeker v. 1.8.0 (Schwengers *et al.*, 2020) and the bacteria RefSeq database (downloaded January 2024) (O'Leary *et al.*, 2016). The reference that was chosen for each ST was defined as:

a) present as a possible reference for the majority of the genomes within that ST

b) the highest average nucleotide identity

c) lowest average MASH distance.

Reads from each ST of interest were aligned to their respective reference genome using Snippy v. 4.6.0 (Seemann, 2015) (Table 5.1). Gubbins v. 3.2.0 (Croucher *et al.*, 2014) was used to remove SNPs attributed to putative recombination from the alignments. The pairwise SNP differences were calculated using snp-dists v.0.7 (Seemann *et al.*, 2021).

Table 5.1: Reference sequence type used to determine pairwise SNP

Sequence type	Genbank Accession assembly ID
ST10	GCF_900636145.1
ST101	GCF_001900815.1
ST1079	GCF_000468515.1
ST1193	GCF_019443685.1
ST12	GCF_010706435.1
ST131	GCF_026651285.1
ST141	GCF_016889685.1
ST164	GCF_020683245.1
ST1800	GCF_020683145.1
ST2163	GCF_904842665.1
ST372	GCF_023658365.1

The threshold to determine the closely related STs within the same sample was five SNPs, as this is the UKHSA threshold for an outbreak of Shiga toxin producing *E. coli* O157 (Mikhail *et al.*, 2018). Genomes with ≤ five SNPs that had the same combination of AMR determinants were defined as putative clones and the presence of putative *E. coli* clones within different hosts is defined as a co-occurrence.

Genomes that were the same ST, had the same combination of antimicrobial resistance determinants, and were ≤ five pairwise non-recombinant SNPs different were labelled as identical. A single representative of each identical group was chosen to represent a putative clonal *E. coli* pair.

FimH alleles were determined using Fimtyper v.1.0 under the Center for Genomic Epidemiology (CGE) server (Center for Genomic Epidemiology; accessed 25/06/2025) (Roer *et al.*, 2017).

5.3.7 Genome wide association analysis

The genome wide association analyses (GWAS) were performed using a unitig and gene presence-absence approach to identify genes associated with host using Pyseer v. 1.3.10 (Lees *et al.*, 2018a). The GWAS was performed on this dataset using two approaches:

1) **Unitig approach**

Unitigs were called from assemblies using unitig-caller v.1.3.0 (Holley and Melsted, 2019). Unitigs associated with the host phenotype (Human or Dog) were identified using a linear mixed model (LMM) approach. The –lmm mode was used with the –distances parameter and a distance matrix estimated by Phylonium v.1.6 to correct for population structure. A covariate matrix consisting of household information was used as input for the –covariates parameter and accounted for household as a potential confounding factor.

2) **Gene presence and absence approach**

Roary v. 3.12.0 (Page *et al.*, 2015) was used to form a gene presence-absence matrix using previously stated parameters. For each gene in the accessory genome, a logistic regression model was used to investigate associations between their presence and host (Human or Dog) using the -lmm approach. To correct for population structure, patristic distances were extracted using the Pyseer v.1.3.10 phylogeny_distance.py script; the distance matrix was used as input for the –distances parameter. The maximum number of dimensions was calculated using the Pyseer v.1.3.10 scree_plot_pyseer command, and the number of dimensions was selected as three using the –max-dimensions parameter according to developer recommendations. A covariate matrix consisting of household information was used as input for the –covariates parameter.

The threshold for significantly associated unitigs and genes was determined for both approaches. The Pyseer –output-patterns parameter was used to generate separate files with hashes corresponding to the unique presence and absence pattern of each unitig and each gene. Using the Pyseer v.1.3.10 count_patterns.py script, the number of unique patterns was calculated for both approaches. A Bonferroni correction was applied to determine the significance threshold for both unitigs and genes, which was calculated separately for the two approaches by dividing 0.05 by the respective total number of unique patterns. Unitigs and genes are considered significantly associated if their respective p-values are lower than the respective Bonferroni-corrected significance threshold.

To assess the confidence in the GWAS results a permutation test was used. Both the unitig and the gene presence and absence approach were applied to the same 10 datasets where the host phenotype (human and dog) associated with each genome was randomly shuffled. The host phenotype associated with each genome was randomised with replacement using the “shuf” command in Linux and the number of genomes was kept consistent. The randomised datasets were analysed using the same parameters as the original analysis.

The candidate hypothetical proteins classified as significantly associated using both approaches were queried against the NCBI nr database [date accessed: 19/09/2024] using BLASTP v.2.16.1+ (Altschul *et al.*, 1997) with default parameters.

5.3.8 Putative plasmid contigs

Putative plasmid contigs were predicted using RFplasmid v.1.0 using the *E. coli* species model. Contigs with a plasmid vote ≥ 0.6 were extracted from assemblies. AMR genes found on these putative plasmid contigs were identified using Abricate v.0.7 using default settings with the Resfinder database using the July 16th 2018 database.

5.4 Results

From the initial 56 households, 10 households were excluded as the dogs were not healthy. Additionally, two households only had faecal samples from one host species and were therefore not analysed. A total of 44 UK households with at least one human and at least one companion dog living in close contact were selected for this study.

In total, 24 households were termed “complete” and had at least one human (n = 31) and at least one dog (n = 27) where at least one *E. coli* colony was recovered (Figure 5.1). Only *E. coli* from complete households were used for further analysis. All households had completed human and dog behavioural data, however one dog had no case enrolment data. All human and dog samples were taken from the first timepoint (T1/T0), except for household 11 and 12 where only samples from timepoint 3 from both human and dog were available. *E. coli* was recovered from 58/92 (63.0%) of all samples. *E. coli* was culturable from 43/54 (79.6%) human samples and 31/48 (64.6%) of dog samples. Only *E. coli* from complete households were used for downstream analysis.

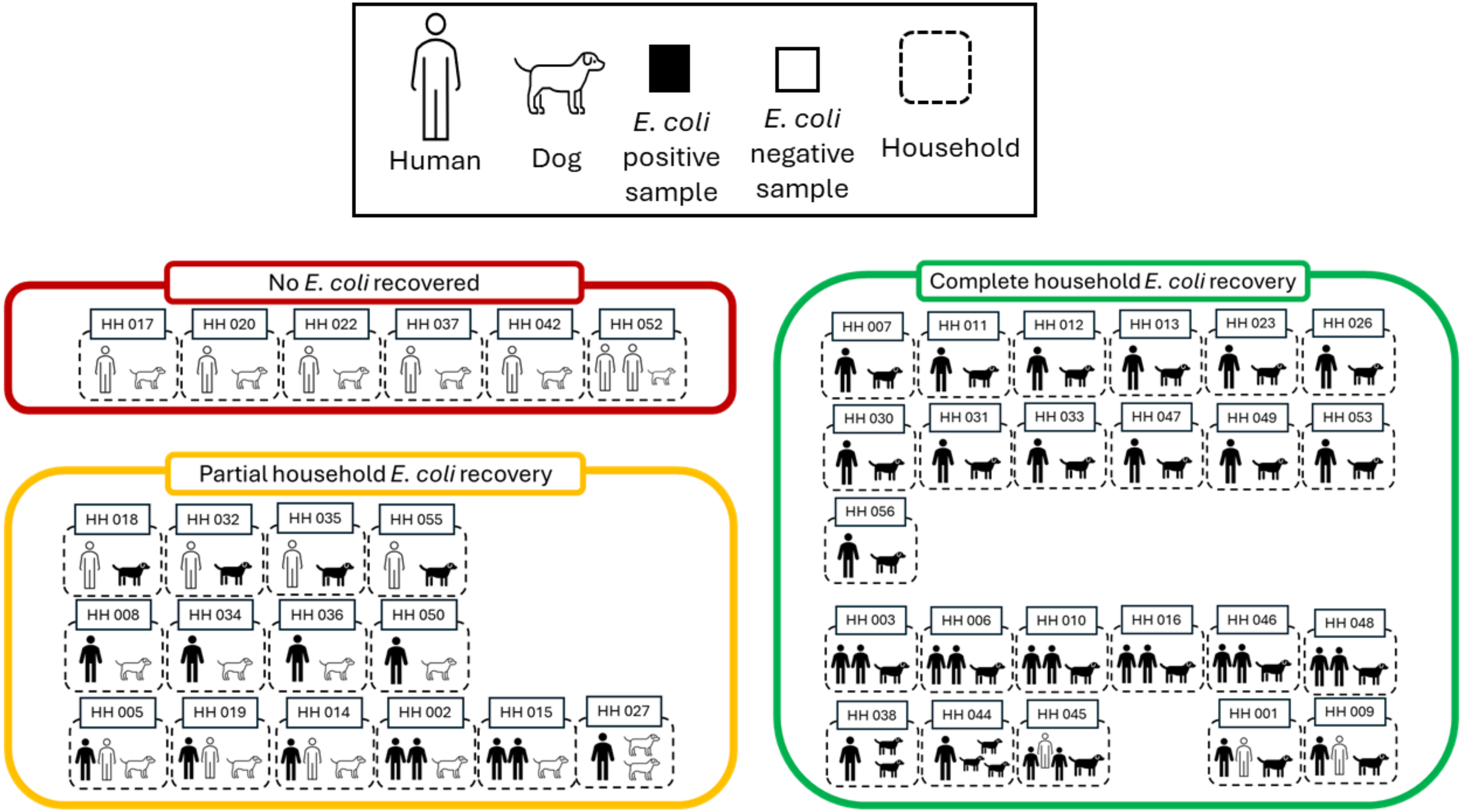


Figure 5.1: *E. coli* recovery from faecal samples from 44 dog-owning households, separated by absence, partial presence and complete households, where at least one human and one dog had *E. coli* recovered from their previously collected faecal samples. Households are denoted by dotted lines, and hosts with *E. coli* positive faecal samples are represented by the presence of colour.

5.4.1 Metadata analysis

All humans from complete households were at least 18 years old, with 67.7% (n = 21) above 40 years old. All companion dogs from complete households had metadata and signalment except for one dog which had no age, weight and sex data available. The dog ages ranged from eight months to 15 years and nine months (median = seven years and nine months). Approximate weights ranged from 5kg to 35kg (median = 14.5kg), 50.0% (n = 13) were male, 50.0% (n = 13) were female.

Only three companion dogs were documented to have been given antibiotics in the last 12 months, and none used antibiotics in the last three months. Additionally, four companion dogs were hospitalised or had surgery in the last 12 months. All dogs from complete households had access to inside the home, and all dogs ate commercial foods, each uniquely supplemented with treats, table scraps, home-made meals, and raw meat treats. In this dataset, five households did not have *E. coli* present in either humans, or companion dogs. Another 15 households had *E. coli* present in only one host species.

5.4.2 Phylogroups and STs

In total, 217 *E. coli* genomes were included in this study and were classified into four predominant phylogroups - B2 (37.8%), B1 (23.5%), A (18.4%), and D (14.3%) (Figure 5.2; Appendix 7). Phylogroup F, cryptic clade III and cryptic clade V were also identified in this dataset. The four predominant phylogroups contained *E. coli* genomes from both host species. This dataset was classified into 53 known STs and one novel ST (Appendix 7, 3), the most common classification being ST131 (10.6%, n genomes = 23). Individual hosts (n humans = 12; n dogs = 12) harboured multiple different STs. A maximum of four STs were recovered from individual humans, whereas a maximum of three STs were recovered from individual dogs (Table 5.2). Within this study, nine STs were found from both humans and dogs – ST2163, ST101, ST12, ST10, ST372, ST1079, ST131, ST1800, and ST164 (Figure 5.3).

Figure 5.2: Phylonium midpoint rooted phylogenetic tree of the collection of 217 *E. coli* genomes isolated from humans and their companion dogs from 24 households, with a tree scale estimating 0.003 nucleotide substitutions per site. The host species (left column), phylogroup (middle column) and household (right column).

Table 5.2: Unique *E. coli* sequence types (STs) identified in faecal samples from humans and their companion dogs from 24 households.

Household	Host	Number of isolates taken	Unique ST
HH 001	D1	4	961
	H1	4	8125
HH 003	D1	4	1442, 978
	H1	5	69
	H2	4	3759
HH 006	D1	4	3576
	H1	4	141
	H2	4	141, 635
HH 007	D1	5	642, Novel6
	H1	4	141
HH 009	D1	4	8573, 720
	H1	4	38
HH 010	D1	3	2163
	H1	4	59, 4429
	H2	1	362
HH 011	D1	1	101
	H1	4	906, 1079
HH 012	D1	4	2538, 589, 5254
	H1	4	710
HH 013	D1	6	399, 7075
	H1	6	95

Table 5.2: Unique *E. coli* sequence types (STs) identified in faecal samples from humans and their companion dogs from 24 households.

Household	Host	Number of isolates taken	Unique ST
HH 016	D1	4	12
	H1	4	9265
	H2	5	12, 73
HH 023	D1	1	6164
	H1	3	10, 372
HH 026	D1	3	101, 349
	H1	6	131
HH 030	D1	3	7345, 7347
	H1	2	372, 404
HH 031	D1	4	10
	H1	5	5295
HH 033	D1	3	12
	H1	4	5345, 95
HH 038	D1	4	106
	D2	4	2559, 101
	H1	4	10, 70
HH 044	D1	4	372, 1079
	D2	4	10
	D3	4	1079, 164
	H1	5	542, 131, 164, 2163
HH 045	D1	4	7236

Table 5.2: Unique *E. coli* sequence types (STs) identified in faecal samples from humans and their companion dogs from 24 households.

Household	Host	Number of isolates taken	Unique ST
	H1	4	131
	H3	2	101, 1800
HH 046	D1	4	942, 131
	H1	4	34
	H2	2	69
HH 047	D1	4	101
	H1	3	131
HH 048	D1	2	131
	H1	2	1193, 131
	H2	4	1193
HH 049	D1	3	1800
	H1	4	73
HH 053	D1	4	5415
	H1	4	58
HH 056	D1	4	131, 372
	H1	4	131, 963

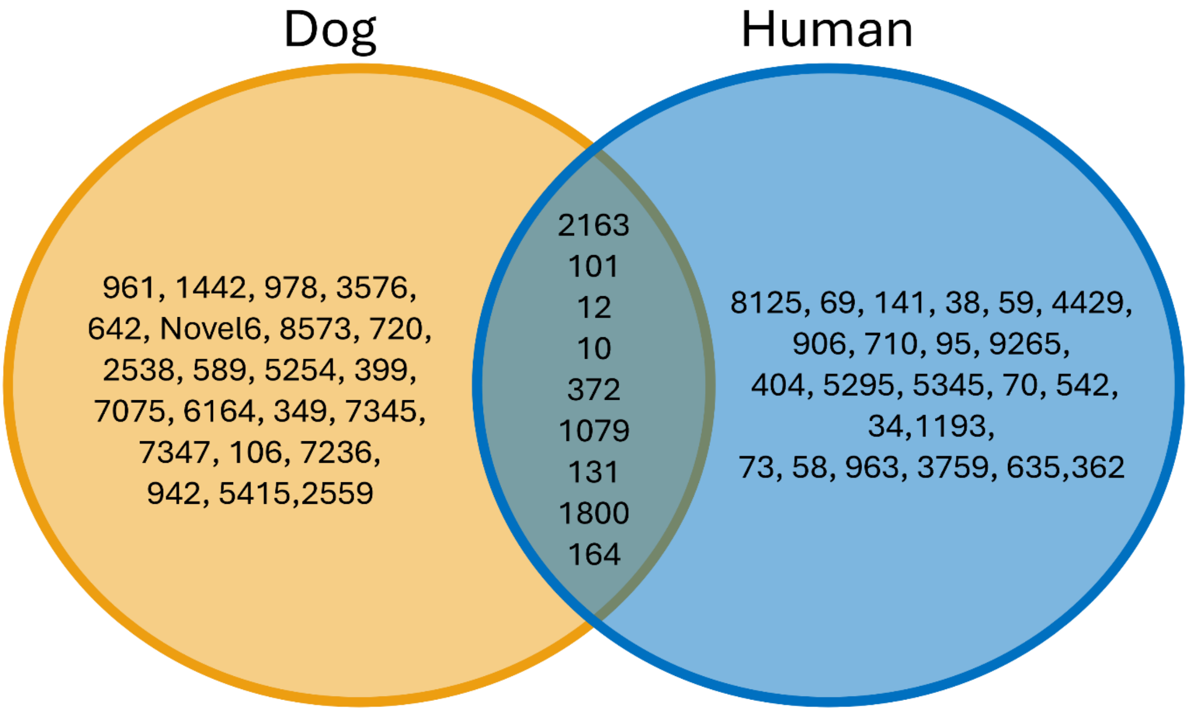


Figure 5.3: Venn diagram of all multilocus sequence types from the 217 *E. coli* genomes isolated humans and their companion dogs from 24 households, separated by host species.

5.4.3 Genome wide association study

The Bonferroni-corrected significance threshold for unitigs was 7.45 x 10^{-8}. Unitig GWAS analysis identified 142 unitigs associated with host species. These unitigs mapped to 45 genes and contig IDs. In this analysis, 12 contig IDs associated with humans and 14 contig IDs associated with dogs did not directly map to an annotated gene (Appendix 8).

The Bonferroni-corrected significance threshold for the gene presence and absence approach was 7.53 x 10^{-6}. The gene presence and absence approach identified 46 clusters of orthologous genes associated with human hosts and 68 associated with dog hosts (Appendix 9).

Using results from both the gene presence and absence approach and the unitig approach, one gene was identified as significantly associated with humans using both methods. The human associated gene was a hypothetical protein, which was identified as the gene ID EC23D1-016-1_04329|TraQ and mapped onto the hypothetical protein group 4813 (Figure 5.4). The datasets where the host phenotype (human vs dog) associated with each genome were randomised had no significantly associated unitigs or genes.

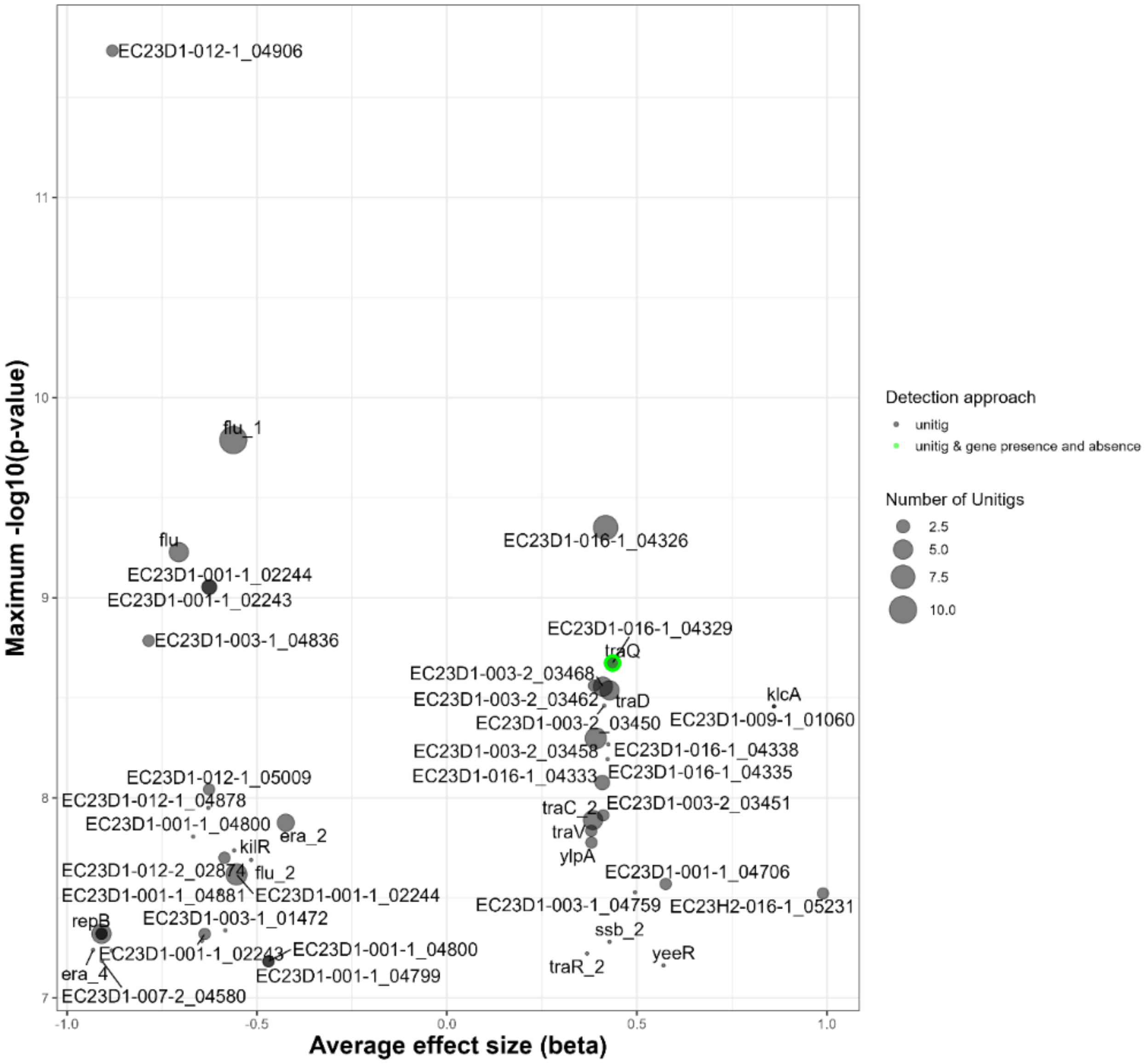


Figure 5.4: The genome wide association showing genes and hypothetical genes that unitigs associated with host species mapped to. The X axis indicates the average effect size (beta) value, where a higher beta value indicates an association to humans, and a lower beta value indicates an association to dogs. The size of the bubble indicates the number of significant unitigs (labelled k-mers) mapping onto the gene. Genes that are significantly associated with a host using both the unitig and pangenome COG presence and absence approach are coloured green.

BLASTn analysis identified a 100% query cover and 100% percent identity (E value = 1e-130) similarity with a type-F conjugative transfer system pilin assembly thiol-disulphide isomerase gene (*TrbB*) (accession number: WP_000059831.1). The human associated hypothetical protein was found in 47 *E. coli* genomes from humans (n genomes = 41; n samples = 13) and dogs (n genomes = 6; n samples = 5). *E. coli* genomes with the group 4813 gene were found in phylogroups A, B2 and D (figure 5.5).

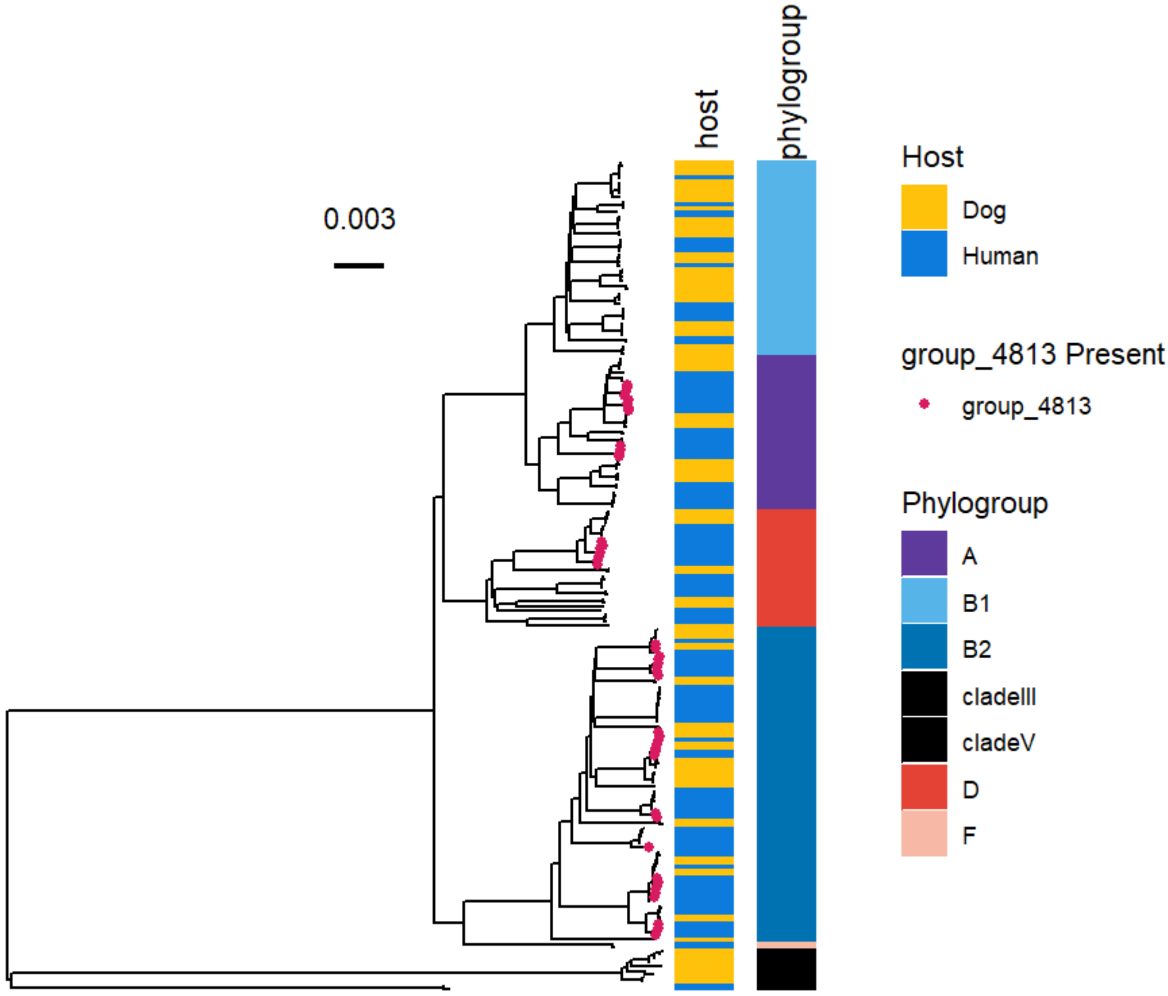


Figure 5.5: Phylonium midpoint rooted phylogenetic tree of the collection of 217 *E. coli* genomes isolated humans and their companion dogs from 24 households, with a tree scale estimating 0.003 nucleotide substitutions per site. The host species (left column), phylogroup (right column) are labelled. Genomes with the human associated hypothetical protein encoding gene, group 4813, are identified by red dots at the end of the tree tips.

5.4.4 Antimicrobial resistance determinants and putative ExPEC

In this chapter, 48.8% of the 217 *E. coli* genomes (n genomes = 106) contained at least one AMR determinant (Appendix 5, 6, 10). Six samples contained genomes that belonged to the same ST but had different AMR genotypes. A maximum of three unique AMR genotypes were identified in one human sample that had four ST710 *E. coli*. A maximum of two unique AMR genotypes were identified in *E. coli* from two dog samples where more than one genome was classified as the same ST (Table 5.3).

Four samples contained *E. coli* genomes assigned to the same ST with differing AMR genotypes, but identical plasmid replicon profiles. EC23H1-013 (ST95) had two AMR genotypes, but four different combinations of plasmid replicons (Table 5.3). Sample EC23H2-006 (ST141) had a genome had no AMR determinants detected and another genome with *bla*TEM-1B, *tet(B),* and four extra plasmid replicons relative to the other ST141 genome (Table 5.3).

Table 5.3: Table of samples displaying *E. coli* from the same sample classified as the same ST with different AMR genotypes and plasmid replicons.

Sample	ST	AMR genotype	Plasmid replicons
EC23D1-031	ST10	No AMR determinants	ColRNAI, IncFIB_AP001918, IncFII_pHN7A8
		23S r.2032T>C	
EC23D1-033	ST12	gyrA p.D87G	No plasmid replicons detected
		No AMR determinants	
EC23H1-012	ST710	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM_1B, cmlA1, dfrA12, floR, sul2, sul3, tet_A	IncFIB_K, IncFII_pCoo, IncI1_I_Alpha
		aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM_1B, dfrA12, floR, sul2, sul3, tet_A	

		aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM_1B, cmlA1, dfrA12, floR, sul3, tet_A	
EC23H1-013	ST95	aph_3____Ib, aph_6__Id, sul2, tet_A	Col_MG828, Col8282, IncB_O_K_Z, IncFIB_AP001918, IncFII, p0111
			Col_MG828, IncFIB_AP001918, IncFII, p0111
		No AMR determinants	IncFIB_AP001918, IncFII, p0111
			IncB_O_K_Z, IncFIB_AP001918, IncFII, p0111
EC23H1-056	ST963	No AMR determinants	IncFIA_HI1, IncFIB_pB171, IncFII_pHN7A8
		23S r.2032G>C	
EC23H2-006	ST141	No AMR determinants	Col_MG828
		blaTEM_1B, tet_B	Col_MG828, Col156, IncFIA, IncFIA_pBK30683, IncFIB_pB171, IncFII_pRSB107

Within this collection, 38 (17.5%) *E. coli* genomes from humans (31 genomes from 13 samples) and dogs (seven genomes from three samples = 7) had putative multi-drug resistance genotypes.

Virulence gene analysis revealed no *stx* genes in this collection, however the *eae* fimbriae gene was identified in two *E. coli* genomes from the same dog sample. 19 genomes across eight samples (n human samples = 6; n dog samples = 2) were classified as putative ExPEC. These genomes belonged to ST95 (n = 8), ST372 (n = 5), ST9265 (n = 4), and ST73 (n = 2). ST95, ST372 and ST73 are from phylogroup B2, and ST9265 from phylogroup D. All samples containing putative ExPEC were from different households.

5.4.5 co-occurrence within and between households

E. coli genomes classified as the same ST, same AMR genotype, and found within the same sample were removed, resulting in 92 genomes (Appendix 11). The 92 *E. coli* genomes were investigated for putative clonal *E. coli* co-occurrences. Using the UKHSA five SNP threshold for clonal outbreak *E. coli*, there were 18 co-occurrences. Of these 18 co-occurrences, five were within the same household. Of these five, three co-occurrences in HH044 (ST164), HH056 (ST131) and HH048 (ST131) were between different host species.

Two co-occurrences in different humans in the same household were observed in HH006 (ST141) and HH048 (ST1193). There were 13 instances of between-household co-occurrences. Furthermore, eight of these 13 between-household co-occurrences were between-different hosts species (Figure 5.6).

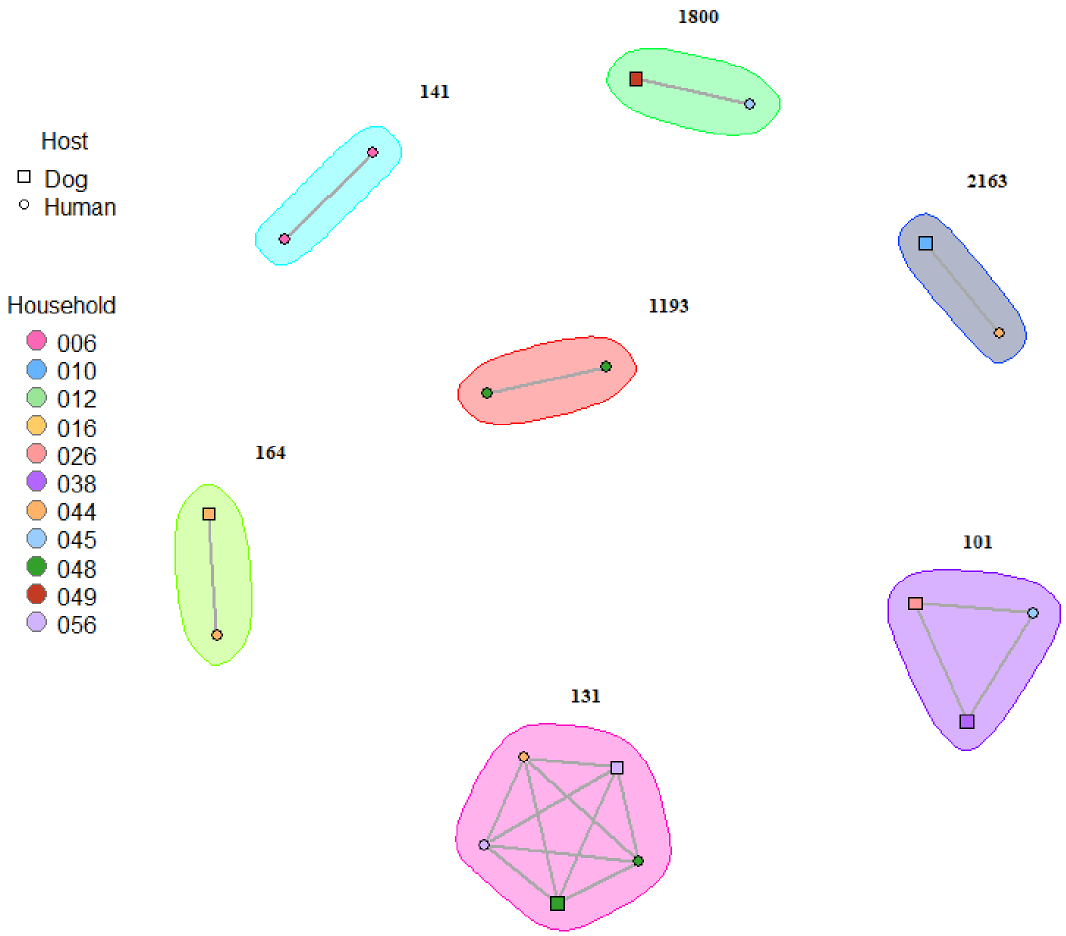


Figure 5.6: Separated networks of putative clonal *E. coli* present in different hosts using the dataset of 92 *E. coli* genomes where replicates were removed from the original 217 *E. coli* isolated humans and their companion dogs from 24 households. The putative clonal *E. coli* co-occurrence between humans (circles) and companion dogs (squares) are indicated by nodes linked by grey lines if there is a five or less single nucleotide polymorphism difference between genomes. Nodes are coloured by the household and grouped by ST.

ST1193-O75:H5 with the *fimH*64 allele co-occurred between two humans in HH048 and harboured 11 identical AMR determinants which were classified into seven different AMR classes (Table 5.4). ST131-O16:H5 with the *fimH*41allele co-occurred in five hosts across HH048, HH056, and HH044 harboured three identical AMR determinants which were classified into three AMR classes (Table 5.4).

Table 5.4: Co-occurring *E. coli* sequence types (STs) identified in faecal samples from humans and their companion dogs with antimicrobial resistance (AMR) determinants. AMR determinants are classified into AMR classes.

Co-occurring ST	AMR determinants	AMR classes
ST1193	aph_3____Ib, aph_6__Id, blaTEM_1B, dfrA17, mph_A, sul2, tet_B, gyrA p.S83L, gyrA p.D87N, par C p.S80I, parE p.L416F	Aminoglycoside, Beta-lactam, Trimethoprim, Macrolide, Sulfonamide, Tetracycline, Quinolone
ST131	aac_3__IId, blaTEM_1B, parE p .I529L	Aminoglycoside, Beta-lactam, Quinolone

Across these 92 genomes, the quinolone resistance group conferred by point mutations was the most abundant AMR determinant in dogs (n = 15/92). The aminoglycoside resistance group was the most abundant AMR determinant in *E. coli* isolated from human samples (n = 36/92). The AMR determinants were classified into nine different antimicrobial drug classes, which included two highest priority critically important antimicrobial drug classes, three critically important antimicrobial drug classes, and four highly important drug classes (Figure 5.7).

E. coli from human samples had more AMR determinants (median = 1, IQR = 3) than *E. coli* isolated from dog samples (median = 0, IQR= 1). A Wilcoxon Rank Sum Test showed a statistically significant difference in number of unique AMR determinants between *E. coli* isolated from dogs and humans (p = 0.01).

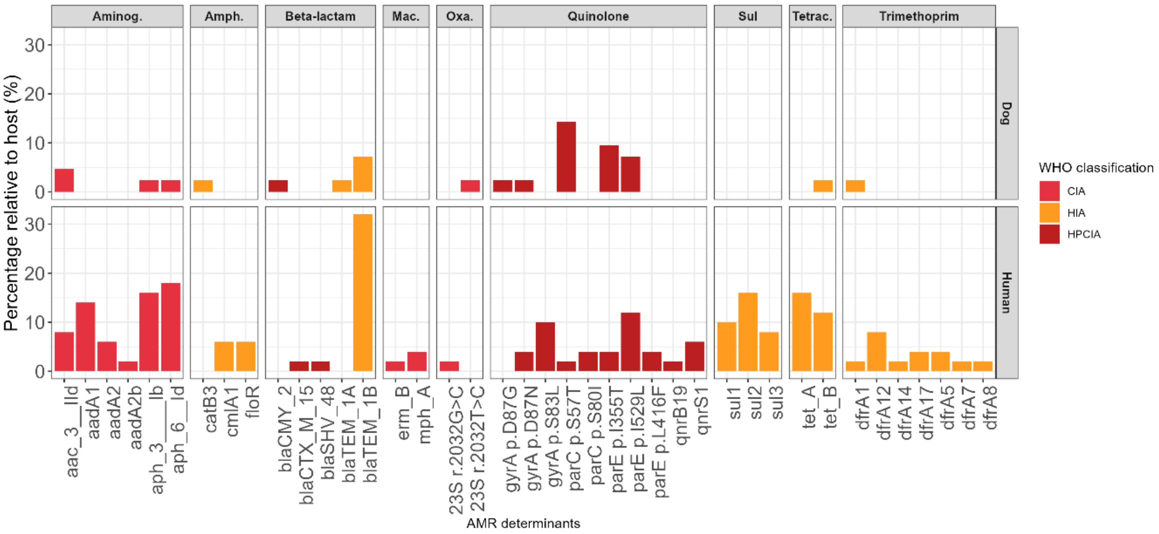


Figure 5.7: The distribution and percentage of antimicrobial resistance (AMR) determinants across 92 *E. coli* genomes (replicates were removed on the basis of the same ST, same AMR genotype, and found within the same sample). AMR determinants are classified into their respective antimicrobial class: aminoglycosides (Aminog.), amphenicols (Amph.), beta-lactams, oxazolidinones (Oxa.), macrolides (Mac.), quinolones, sulphonamides (Sul.), tetracyclines (Tetrac.), and trimethoprim. Antimicrobial classes are classified as Highly Important (HIA), Critically Important (CIA), and Highest Priority Critically Important (HPCIA) according to the WHO Medically Important Antimicrobials List (WHO MIA List). *E. coli* from humans (n = 50) and companion dogs (n = 42) are separated.

5.4.6 Putative plasmid contigs

The two unitig sequences associated with the human associated hypothetical gene, group 4813, were found within 22 plasmid contigs from 22 different genomes belonging to six STs (Table 5.5). Of these 22 genomes, 18 contained the hypothetical gene, group 4813. Plasmid contigs with the human associated unitigs present were found in five ST12 genomes from the same household from both the human and companion dog. One of these ST12 genomes from a human sample had a 68,569bp putative plasmid contig with the human associated unitigs, as well as antimicrobial resistance genes *tet*B and *bla*$_{\text{TEM-1B}}$.

Table 5.5: *E. coli* genomes within this dataset with human associated unitigs present in plasmid contigs separated by sequence type.

Sequence type	***E. coli* genomes from dogs with human associated**	***E. coli* genomes from humans with human**

	unitigs present in plasmid contigs	**associated unitigs present in plasmid contigs**
ST12	3	2
ST131	0	7
ST34	0	1
ST69	0	1
ST73	0	3
ST8125	0	1

5.5 Discussion

5.5.1 Limited *E. coli* co-occurrence

The limited within-household co-occurrences in our study (3/24 households) is consistent with previous lower resolution macro-restriction PFGE approaches, which reported cross-species co-occurrences in 4-9.8% of households (4.00%, 1/25 households; 8.82%, 3/34 households; 9.84%, 6/61 households) (Stenske *et al.*, 2009, Naziri *et al.*, 2016, Harada *et al.*, 2012). Similarly, previous studies found limited human-dog cross-species clonal ESBL/AmpC-producing *E. coli* co-occurrences within 0.909 – 8.7% of households (0.909%, 5/550 households; 8.70%, 2/23 households) using selective culture and WGS (van den Bunt *et al.*, 2020, Toombs-Ruane *et al.*, 2020). A prospective longitudinal WGS study identified two animal-human pairs (2/59 households) with co-occurring ESBL-producing *E. coli* (Menezes *et al.*, 2023). In that longitudinal study, a dog showed long-term ESBL-producing *E. coli* carriage and subsequent presence within the owner suggesting potential transmission (Menezes *et al.*, 2023). According to our work and previous work, clonal *E. coli* co-occurrence between dogs and humans is limited.

Within clinical settings, longitudinal data from individual patients were used to establish a 17 SNP threshold indicating putative transmission (Ludden *et al.*, 2021). As this chapter describes a cross-sectional study, it is difficult to establish SNP accumulation rates and SNP thresholds for transmission. Strong evidence for transmission between humans and dogs is lacking in this chapter; this would require detection of the same *E. coli* lineage in each host across multiple timepoints.

It is also important to take multiple *E. coli* isolates when investigating putative clonal co-occurrence between hosts, as this chapter identified that hosts could have multiple different STs. Previous work has also shown that up to 10 STs were identified when 16 colonies were tested in a human stool sample (Stoesser *et al.*, 2015) and up to eight STs within a human when up to 15 colonies were tested (Ludden *et al.*, 2021). This within-host within-ST *E. coli* SNP variation has also been observed for both humans (Ludden *et al.*, 2021) and dogs (Hayer *et al.*, 2023).

The ST variation may be because these STs are long term gut colonisers of the host, where an *E. coli* lineage colonised the host, and then accumulated SNPs within the gut. Alternatively, different lineages of the same *E. coli* ST, could be introduced to the host through an external environment or different host. As shown in chapter 4 of this thesis, food may be a possible route of exposure to diverse *E. coli* belonging to the same ST. Furthermore, shared food has also been highlighted as a risk factor for *E. coli* similarity between owners and dogs (Naziri *et al.*, 2016). There is also the possibility that the diverse *E. coli* in the gut could arise from both the introduction of *E. coli* from an external source and SNP accumulation. Other routes of transmission could be equally important and should be investigated, highlighting the complexity of potential transmission between hosts in wider society.

5.5.2 A stringent genome wide association study highlights only one gene associated with human hosts

A stringent approach using a population-corrected genome wide association study showed that the type-F conjugative transfer system pilin assembly thiol-disulphide isomerase (*TrbB*) gene was associated with *E. coli* isolated from humans. The thiol-disulphide isomerase genes mediate disulphide bond formations when generating nascent proteins (Bekendam and Flaumenhaft, 2019). These thiol-disulphide isomerases are typically used in post-translation modification with the *TraQ* (prepilin peptidase) to convert pre-pilin into pilin (Dupuy *et al.*, 2013). This pilin can polymerise and form a five strand helical structure – the F-pilus (Costa *et al.*, 2016). This F-pilus is useful for conjugation and the transport of plasmid DNA between Gram-negative bacteria (Patkowski *et al.*, 2023). This gene is typically found on plasmids (Patkowski *et al.*, 2023) and is supported by the presence of these human-associated untigs on plasmid contigs within this study. The conjugation process is central for the spread and distribution of mobile AMR genes, and can be facilitated by F-pili (Patkowski *et al.*, 2023).

The *bla*$_{\text{TEM-1B}}$ in this chapter was identified on a plasmid contig that also contained the human associated unitig sequence. The *bla*$_{\text{TEM-1B}}$ AMR determinant can confer resistances to penicillin and extended spectrum beta-lactams when expressed, and can be both chromosomal or mobilisable (Farooq *et al.*, 2025). This human-associated gene (in this study) was identified across multiple different phylogroups, STs, within both humans and dogs, and identified on putative plasmid contigs, suggesting possible acquisition through horizontal transfer.

However, the association of this gene to humans should be interpreted with caution. There may be a plasmid-mediated co-selection process occurring, where antibiotic use may be indirectly selecting for this gene because of the linked AMR determinants. Additionally, the association of this gene to humans also does not necessarily mean that this is a result of adaptation. Furthermore, as this dataset has a limited sample size, further work is required to validate this result and identify the generalisability of this finding on a larger scale.

5.5.3 *E. coli* from humans in this dataset is more likely to carry AMR determinants

The *E. coli* isolated from humans harboured more AMR determinants than from dogs. Our study suggests that the companion dogs in this study may not necessarily be a major AMR risk to their owners. Alternatively, the owners may harbour *E. coli* with AMR and present a risk to their pets. This possible human-mediated risk corroborates with previous work where an MDR and virulent *E. coli* subclone carriage was identified in human populations (Rodrigues *et al.*, 2016), and the subsequent emergence of the same MDR and virulent *E. coli* clonal group identified in companion animals (Belas *et al.*, 2018).

AMR genotype variation within the same ST in the same sample indicates that the same ST does not indicate the same AMR risk. The presence of multiple plasmid replicon profiles across the same sample also suggests some role of mobile genetic elements in shaping the *E. coli* population observed within humans and dogs. Differences in plasmid replicon profiles occurring with differing AMR genes within the same ST may be due to biological processes like horizontal gene transfer. However, this variation may also be due to short-read sequencing errors and fragmented assemblies around repetitive plasmid regions which fail to resolve plasmids and fail to detect mobile genetic elements (Juraschek *et al.*, 2021). Conversely, the same AMR genotype with differing plasmid replicon profiles suggests that the AMR determinants in this work may not necessarily be plasmid mediated. This is evident by the point mutations in ST10 found in sample EC23D1-031 which shows putative AMR can arise independently of plasmids and mobile genetic elements.

Within the same sample, the same ST could have different AMR genotypes and The overall presence of putative ExPEC in this chapter was limited. While the cross-sectional design prevents inferring transmission, if there were transmission, the low prevalence of ExPEC-positive samples suggests the overall health risks are low (n human samples = 6; n dog samples = 2). Virulence gene analysis revealed no *stx* genes in this collection, however the *eae* fimbriae gene was identified in two *E. coli* genomes from the same dog sample. Previous work suggests that this gene has been found in *E. coli* from healthy dogs, and domestic animals have been highlighted as potential reservoirs for *E. coli* with this gene (Holland *et al.*, 1999, Krause *et al.*, 2005). Furthermore, studies on whether dogs or humans are reservoirs of ExPEC which infect the other species are conflicting (Damborg *et al.*, 2009, Elankumaran *et al.*, 2022). However, the samples in this chapter showed no evidence of putative ExPEC co-occurrence between dogs and humans.

Putative MDR ST131 and ST1193 co-occurred between different hosts in this dataset. The ST131-O16:H5 with the *fimH*41 allele lineage identified in this dataset co-occurred across host species and between households, falls within the previously described ST131-A clade (Matsumura *et al.*, 2017), and has been identified in both human clinical cases and companion animals (Dahbi *et al.*, 2014, Kidsley *et al.*, 2020b). In addition, ST1193-O75:H5 with the *fimH*64 allele has been highlighted as an emerging globally disseminated MDR ST (Valenza *et al.*, 2019, Johnson *et al.*, 2019). In our study, ST1193-O75:H5 with the *fimH*64 allele co-occurred between two humans within the same household, but previous work has shown presence in both humans and dogs (Johnson *et al.*, 2019).

Both ST131 and ST1193 have been highlighted as high-risk clones, which can cause urinary and bloodstream infections (Pitout *et al.*, 2022). High-risk clones are defined as globally spread, possessing various AMR determinants, with abilities to persist and colonise a host for at least six months, spread between hosts, show increased pathogenicity, and cause severe/recurring infections (Pitout *et al.*, 2022). As this is a cross-sectional genomic study, the persistence and ability to cause severe/recurring infections cannot be determined. Our findings suggest both owners and pets can potentially aid in the dissemination of ST131-O16:H5 and ST1193-O75:H5 clones; it is therefore important to consider both humans and companion animals within household units for assessing potential dissemination of AMR *E. coli*.

5.5.4 *E. coli* was not recoverable from all faecal samples

The *E. coli* faecal sample recovery rate within this chapter may be affected by the freeze-thaw process during long-term storage. This freeze-thaw process may injure bacterial cells and reduce the number of culturable *E. coli* (Ternent *et al.*, 2004). Moreover, the TBX agar growth medium used, cannot detect around 5% of *E. coli* that are glucuronidase negative (Hansen and Yourassowsky, 1984) like *E. coli* O157 (Ratnam *et al.*, 1988). This may limit our observations on the diversity of *E. coli* within these faecal samples. Overall, *E. coli* carriage rates are variable, ranging from 88.0% - 95% in humans (Bailey *et al.*, 2010, Adenipekun *et al.*, 2016). When comparing different culture-approach studies in healthy dogs, *E. coli* presence in faeces and rectal samples have been found to vary between 10.9% to 100% (Buranasinsup *et al.*, 2023, Schmidt *et al.*, 2015, Karahutová *et al.*, 2021, Menezes *et al.*, 2023). Differences in *E. coli* recovery rates between these studies and this chapter is likely due to differences in the stool sample collection, whether antimicrobials were used and the culture method.

5.6 Conclusion

In this work, there were limited instances of putative clonal *E. coli* co-occurrence between humans and their pet dogs using a WGS approach without selective enrichment for AMR. Although co-occurrence within households was limited, the co-occurrence of clonal *E. coli* between humans and dogs from different households indicates the complexity of potential sharing between hosts in wider society.

One genomic feature associated with *E. coli* from human hosts was identified in this dataset with possible roles in conjugation, although it is necessary to investigate whether or not this is applicable to other datasets. As this genomic feature was also found on putative plasmid contigs, further work can explore whether plasmids are found within these genomes.

This study contributes to the growing body of evidence showing that clonal *E. coli* with potential to harbour AMR determinants can be present in both humans and their cohabiting animals. While the directionality of *E. coli* and AMR transmission between humans and pets cannot be inferred in this study, it is evident that both hosts can contribute to the dissemination and spread of clinically relevant *E. coli* clones of concern.

A more uniform approach for assessing AMR in *E. coli* from both healthy humans and companion dogs ensures more accurate estimation of AMR and managing risk. It is important to take all hosts within the household unit into account when identifying and implementing interventions for reducing the spread of clinically important AMR *E. coli*. However, other routes of transmission could be equally important and should be investigated. This highlights the complexity of potential transmission between hosts in wider society.

Chapter 6: General Discussion

Both retail foods and the household environment are examples of complex environments for the dissemination and potential circulation of different *E. coli* lineages. The majority of *E. coli* types are commensal with potential as opportunistic pathogens, with a subset of the species being pathogenic (Leimbach *et al.*, 2013, Geurtsen *et al.*, 2022). This is a health issue as *E. coli* can acquire or develop AMR, making treatments difficult (Murray *et al.*, 2022). WGS has proven useful for epidemiological investigations of outbreaks relating to foods, humans and animals, and has significant applications to UK government policies for public health. When evaluating the literature and current UK surveillance systems for *E. coli* in foods, humans and animals, there has been a primary focus on pathogenic lineages that cause immediate gastrointestinal disease, lineages causing opportunistic infections, and specific AMR circulating in *E. coli*. A selective approach for identifying specific lineages would be appropriate for assessing immediate public health risks but fails to acknowledge and appreciate the diversity of *E. coli* within different niches and the variation in risk these environments present to humans. This thesis addresses gaps in knowledge surrounding *E. coli* evolutionary history and diversity on retail foods and within hosts co-habiting within the same household.

6.1 Discussion of key findings

One of the key findings of this thesis in chapter 4 is that retail foods can act as a fomite for highly diverse lineages of *E. coli*, which may then present a risk of exposure to the consumer. This has implications for food safety policy, where current contamination surveillance approaches do not assess the diversity of *E. coli* present. However, the exposure to diverse populations of *E. coli* from food does not mean that all the strains contaminating the food will become present in the host gastrointestinal tract. It was evident when comparing *E. coli* strains from dogs and humans in chapter 5 that although the household environment provides multiple opportunities for *E. coli* co-occurrence, the observed number of co-occurrence events was low. Both studies were cross-sectional which may limit the observed diversity and co-occurrence. Therefore, to assess whether or not food is a major contributor to the diversity of *E. coli* in the human gut, a multi-isolate longitudinal WGS approach could be taken where humans and the food they are about to consume are cultured for *E. coli*. This has been well-documented for pathogenic lineages of *E. coli* such as STEC, where food can be a fomite (Wilson *et al.*, 2018), but the dynamics of commensal *E. coli* on food and whether they can reach the consumer gut is not well understood. Further food processing at home (i.e., cooking) is likely to reduce the abundance and types of *E. coli* found on food to which humans are exposed. However, the presence of *E. coli* on ready-to-eat fresh leafy greens and cooked prawns in this study suggests there may be a considerable risk of exposure. This work could be used as a starting point for understanding food-human *E. coli* co-occurrences and the effect of cooking and washing on potential transmission.

A longitudinal WGS *E. coli* experiment could also be applied to non-human hosts such as companion dogs to help determine the directionality of the co-occurrences between these host species identified in chapter 5. This is of particular concern with increasingly popular raw meat fed (RMF) diets that may introduce both pathogenic lineages and AMR *E. coli* to companion dogs (Morgan *et al.*, 2024, van Bree *et al.*, 2018). The household *E. coli* work was based on a longitudinal study (Menezes *et al.*, 2023), but due to the low recovery rates for *E. coli* in the stool samples, the work presented in this thesis is cross-sectional, and only co-occurrence could be investigated. Although limited co-occurrence between species and within household was observed, two MDR clones co-occurred between humans and companion dogs. Again, a multi-isolate longitudinal WGS experiment focussing on commensal *E. coli* in humans and dogs, where the food is also monitored and sampled would provide more insights into the human-companion animal *E. coli* sharing interface, whilst also providing a One Health insight into potential exposure routes.

Another key finding was that there was a lack of phylogenetic clustering according to food commodity or host species using the phylogenomic method determined in chapter 3 applied to *E. coli* from different food commodities in chapter 4 and host species in chapter 5. This may be because of temporary contamination on food, or a lineage transiently present within host species, leaving little time for host-specific clusters to emerge. This work captured snapshots of different transient *E. coli*, or alternatively transient *E. coli* may co-occur with established long-term lineages, leading to diverse populations in each niche, which would require intensive sampling to identify. In either case it is difficult to draw conclusions as to which are adapted for long-term colonisation of the niche, and which are adapted for transient survival.

In terms of AMR, this thesis highlights that these diverse *E. coli* populations in different environments may expose hosts (human or animal) to a high number of different AMR determinants. A selective approach where specific antimicrobial resistant *E. coli* is investigated would lead to a reduction in the range of AMR we would observe on food, in humans and in pets. The work within this thesis suggests that the diversity of AMR varies between food commodity and sample, as well as within humans and dogs, as evident by the different AMR profiles. This is important because if AMR were being surveyed within different host populations and environments, then multiple *E. coli* isolates would need to be analysed to ensure a more thorough AMR risk assessment. Moreover, work is required to understand which AMR determinants from food can successfully become present in *E. coli* within a host, and which AMR determinants are shed and potentially disseminated between hosts within a household. A collaborative multi-disciplinary One Health approach is necessary for assessing, preventing, and responding to pathogen spread and AMR. This would be particularly useful for public health policy, as it would provide more clear points of AMR dissemination which intervention strategies could be applied to.

6.2 Future work

Methodologically, the work within this thesis builds upon previous studies in the field by isolating and sequencing multiple isolates per food sample. However, this work also indicates that for most of the food commodities investigated, the number of isolates taken from each food sample (up to four) was likely insufficient for capturing the diversity of *E. coli* in this niche, with the exception of salmon. A similar finding was identified in chapter 5, where up to eight *E. coli* isolates were investigated from humans and dogs. More within-household between-hosts may have been observed if more isolates from each human and dog were investigated.

There are also limitations with all phylogenomic approaches, especially when analysing new datasets with unknown levels of recombination and horizontal gene transfer. Therefore, to fully test the limitations of these phylogenomic approaches, future work could create simulated datasets with increasing levels of recombination and horizontal gene transfer. Similar work by Lees et al. (2018) utilised similar approaches with *Streptococcus pneumoniae*, which created simulated datasets with varying SNPs, short insertions and deletions, and simulation of evolution in intergenic regions (Lees *et al.*, 2018b). Additionally, variation in parameters within the phylogenomic methods, such as inclusion thresholds for percentage of core genes, could also change the final tree.

All genomic analyses undertaken in this thesis utilised short read WGS which may cause difficulty with assembly around highly repetitive regions. MGEs often contain repetitive regions and are thus assembled into multiple contigs. Long read sequencing could be applied to the *E. coli* datasets in chapters 4 and 5 to help identify the presence of MGEs (i.e., plasmid, integrons, insertion sequences, bacteriophage), and which accessory genes are associated with these different MGEs. This would be particularly relevant in chapter 5, which identified a gene involved in conjugation associated with humans, and the context of other genes that co-occur with *trbB*.

The companion dogs investigated within chapter 5 harboured diverse *E. coli* populations which could then be shed, and similar findings have been identified with farm animals (Li *et al.*, 2020, Ahmed *et al.*, 2017). One possibility for the diverse population of *E. coli* found on many meat samples in chapter 4 may start with the host animal. However, this is not the only way contamination can be introduced; this could also happen throughout the food manufacturing process, with cross-contamination events from different hosts occurring, as well as contaminated equipment, or even from workers. An investigation on the complete food manufacturing process from host animals, food processing, packaging and eventual display within retail stores would identify the types of *E. coli* resulting from specific contamination sources and possible interventions that could be implemented to prevent contamination, reducing exposure risk for consumers.

Metagenomics would be a promising addition to a longitudinal multi-niche study investigating host animals and the food processing environment. Short-read metagenomics can be used to examine entire microbial populations at the species level within a niche. However, there are limitations with read length and assembly fragmentation in short-read metagenomics. Long-read metagenomics can overcome these limitations by producing longer, more complete genome assemblies (Kim *et al.*, 2024). For example, a long-read metagenomics workflow has also been proposed to for improved STEC detection and characterisation in irrigation water contamination for food safety contexts (Maguire *et al.*, 2021). A combined short-read and long-read metagenomic approach for investigating retail foods is a promising approach for improving food safety surveillance, although there are still limitations detecting low abundance organisms (Bloomfield *et al.*, 2025).

Short-read metagenomics has also been used to investigate human and pet dog microbiomes, where pet dogs resemble their owners AMR genes, MGEs and microbiome composition (Zhao *et al.*, 2022). A longitudinal 16s rRNA metagenomics study on ESBL-*E. coli* carriage in dogs also applied long-read metagenomics to two dog samples, which provided insight into the locations of AMR genes and integrons (Stege Paul *et al.*, 2023). To build on the work within this thesis, long-read metagenomics could be applied to a longitudinal study to better understand *E. coli* population diversity and transmission dynamics across the human, animal, and food niches.

6.3 Proposed model of *E. coli* diversity

The complexity within all these environments comes from unknown points of contamination or points of exposure that can introduce *E. coli* with different evolutionary histories. It is also important to note that exposure will not necessarily lead to colonisation and subsequent survival within a host; it is likely for a lineage of *E. coli* to successfully contribute to the diversity with a niche, it will likely depend on factors such as: sufficient number of initial contaminating/colonising bacterial cells, tolerance to changes in metabolites present, external stress tolerance, length of exposure time, and number of exposure events. A proposed model (Figure 6.1) for the possible routes of exposure which may be contributing to the *E. coli* diversity in each environment highlights that there are many unknowns that contribute to the end population we observe in our cross-sectional studies. Future work exploring these environments should occur together and have longitudinal aspects to model the diverse *E. coli* communities, allowing us to understand how *E. coli* diversity is created and maintained in these niches and how they can contribute varying levels of direct and indirect health risks to hosts.

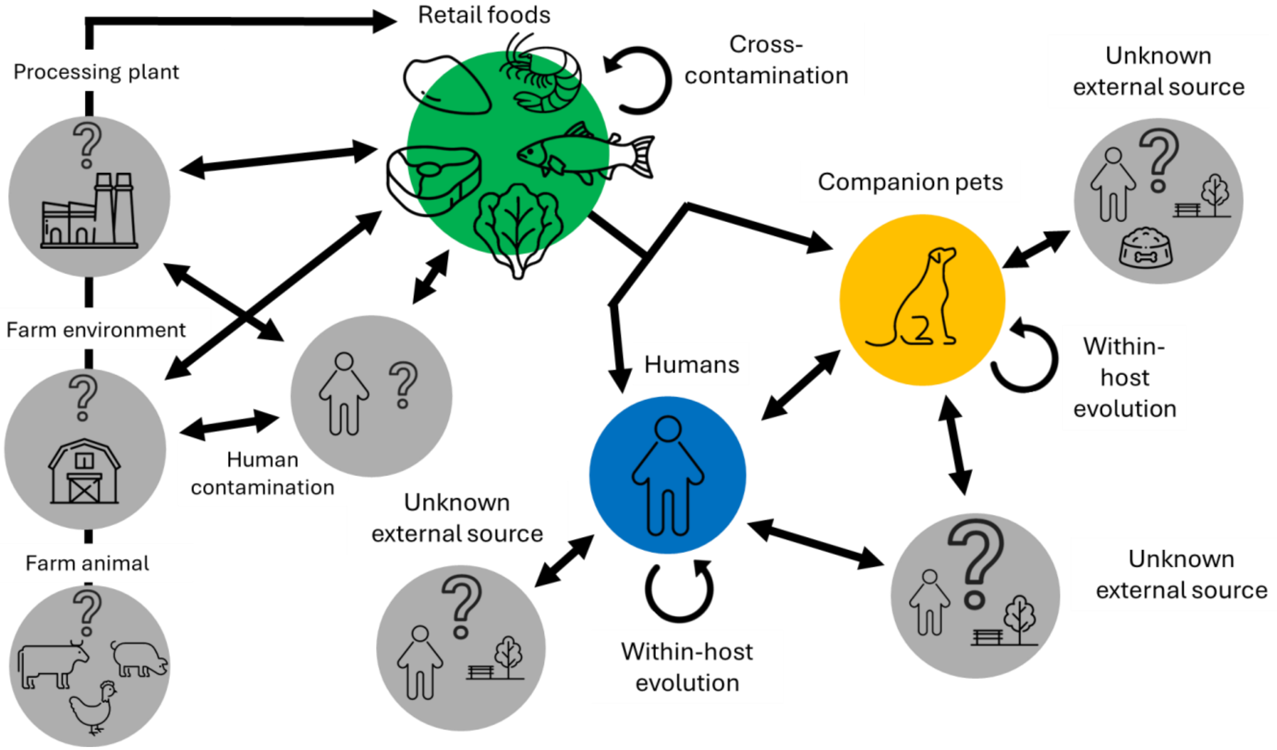


Figure 6.1: A flow diagram indicating the potential contamination and exposure routes that may result in the diverse populations of *E. coli* identified within this thesis. The environments which may be points of exposure are highlighted, with arrows indicating points where different *E. coli* may be introduced. Within host evolution and potential cross-contamination is denoted by a circular arrow and annotated. Work explored in this thesis is highlighted by the coloured groups - the retail food niche is coloured in green; the human and dog niche are coloured in blue and yellow respectively.

6.4 Final remarks

E. coli is present in a wide variety of niches, each with different populations and AMR risks. This work has made significant contributions to understanding *E. coli* populations and diversity on food, in humans and in animals. Appropriate phylogenomic approaches, increased sampling efforts, and WGS are necessary for exploring *E. coli* populations and AMR risk in the three compartments of the One Health triad. The data generated from the comprehensive cross-sectional multi-isolate WGS study of *E. coli* on retail foods in the U.K., and the work investigating *E. coli* co-occurrence between co-habiting humans and dogs, will be useful for informing public health safety. Cross-disciplinary collaborations between policy makers and scientists are necessary for monitoring *E. coli* in different niches and for informing public health policy, with a thorough understanding of the niche connections and microbial ecology. The ecology of *E. coli* is complex, but these niches do not exist in isolation. As we continue to exist with other hosts and within our environment, we are not only responsible for ourselves but for the impact we have on each other.

References

Abram, K., Udaondo, Z., Bleker, C., Wanchai, V., Wassenaar, T.M., Robeson, M.S. and Ussery, D.W. (2021) Mash-based analyses of *Escherichia coli* genomes reveal 14 distinct phylogroups. *Communications Biology,* 4**,** 117. Available at: https://doi.org/10.1038/s42003-020-01626-5

Adenipekun, E.O., Jackson, C.R., Ramadan, H., Iwalokun, B.A., Oyedeji, K.S., Frye, J.G., Barrett, J.B., Hiott, L.M., Woodley, T.A. and Oluwadun, A. (2016) Prevalence and multidrug resistance of *Escherichia coli* from community-acquired infections in Lagos, Nigeria. *Journal of infection in developing countries,* 10**,** 920-931. Available at: https://doi.org/10.3855/jidc.7997

Advisory Committee on the Microbiological Safety of Food. (2020) *ACMSF subgroup on non-proteolytic Clostridium botulinum and vacuum and modified atmosphere packaged foods Final report*. Food Standards Agency. Available: https://acmsf.food.gov.uk/sites/default/files/acm-1322-vpmap-report.pdf [Accessed 11/08/2025].

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A. and Goecks, J. (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res,* 44**,** W3-W10. Available at: https://doi.org/10.1093/nar/gkw343

Ahmed, S., Besser, T.E., Call, D.R., Weissman, S.J., Jones, L.P. and Davis, M.A. (2016) Evaluation of two multi-locus sequence typing schemes for commensal *Escherichia coli* from dairy cattle in Washington State. *Journal of Microbiological Methods,* 124**,** 57-61. Available at: https://doi.org/10.1016/j.mimet.2016.03.008

Ahmed, S., Olsen, J.E. and Herrero-Fresno, A. (2017) The genetic diversity of commensal *Escherichia coli* strains isolated from non-antimicrobial treated pigs varies according to age group. *PLOS ONE,* 12**,** e0178623. Available at: https://doi.org/10.1007/s00248-010-9717-3

Almeida, J.S. (2013) Sequence analysis by iterated maps, a review. *Briefings in Bioinformatics,* 15**,** 369-375. Available at: https://doi.org/10.1093/bib/bbt072

Almeida, J.S., Carriço, J.A., Maretzek, A., Noble, P.A. and Fletcher, M. (2001) Analysis of genomic sequences by Chaos Game Representation. *Bioinformatics,* 17**,** 429-437. Available at: https://doi.org/10.1093/bioinformatics/17.5.429

Alsultan, A. and Elhadi, N. (2022) Evaluation of ERIC-PCR method for determining genetic diversity among *Escherichia coli* isolated from human and retail imported frozen shrimp and beef. *International Journal of Food Contamination,* 9**,** 12. Available at: https://doi.org/10.1186/s40550-022-00098-1

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-3402. Available at: https://doi.org/10.1093/nar/25.17.3389

Álvarez-Ordóñez, A., Alvseike, O., Omer, M.K., Heir, E., Axelsson, L., Holck, A. and Prieto, M. (2013) Heterogeneity in resistance to food-related stresses and biofilm formation ability among verocytotoxigenic *Escherichia coli* strains. *International Journal of Food Microbiology,* 161**,** 220-230. Available at: https://doi.org/10.1016/j.ijfoodmicro.2012.12.008

Anderson, K.L., Casey, R.A., Cooper, B., Upjohn, M.M. and Christley, R.M. (2023) National dog survey: describing UK dog and ownership demographics. *Animals (Basel),* 13**,** 1072. Available at: https://doi.org/10.3390/ani13061072

Andersson, M., Östholm-Balkhed, Å., Fredrikson, M., Holmbom, M., Hällgren, A., Berg, S. and Hanberger, H. (2019) Delay of appropriate antibiotic treatment is associated with high mortality in patients with community-onset sepsis in a Swedish setting. *European Journal of Clinical Microbiology & Infectious Diseases,* 38**,** 1223-1234. Available at: https://doi.org/10.1007/s10096-019-03529-8

Animal and Plant Health Agency. (2024) *General submission form for diagnostic tests*. Available: https://www.gov.uk/government/publications/general-submission-form-for-diagnostic-tests [Accessed 21/03/2025].

Animal and Plant Health Agency. (2025) *Zoonoses and veterinary public health: annual report 2024*. Available: https://www.gov.uk/government/publications/zoonoses-and-veterinary-public-health-disease-surveillance-reports [Accessed 10/06/2025].

Arenas, M., Valiente, G. and Posada, D. (2008) Characterization of Reticulate Networks Based on the Coalescent with Recombination. *Molecular Biology and Evolution,* 25**,** 2517-2520. Available at: https://doi.org/10.1093/molbev/msn291

Austin, B. (2006) The bacterial microflora of fish, revised. *The Scientific World Journal,* 6**,** 931-945. Available at: https://doi.org/10.1100/tsw.2006.181

Avni, E. and Snir, S. (2020) A New Phylogenomic Approach For Quantifying Horizontal Gene Transfer Trends in Prokaryotes. *Scientific Reports,* 10**,** 12425. Available at: https://doi.org/10.1038/s41598-020-62446-5

Aworh, M.K., Thakur, S., Gensler, C., Harrell, E., Harden, L., Fedorka-Cray, P.J. and Jacob, M. (2024) Characteristics of antimicrobial resistance in *Escherichia coli* isolated from retail meat products in North Carolina. *PLOS ONE,* 19**,** e0294099. Available at: https://doi.org/10.1016/j.mimet.2016.03.008

Bailey, J.K., Pinyon, J.L., Anantham, S. and Hall, R.M. (2010) Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *Journal of Medical Microbiology,* 59**,** 1331-1339. Available at: https://doi.org/10.1099/jmm.0.022475-0

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology,* 19**,** 455-477. Available at: https://doi.org/10.1089/cmb.2012.0021

Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E. and Clermont, O. (2018) ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microbial Genomics,* 4**,** e000192. Available at: https://doi.org/10.1099/mgen.0.000192

Begley, M. and Hill, C. (2015) Stress Adaptation in Foodborne Pathogens. *Annual Review of Food Science and Technology,* 6**,** 191-210. Available at: https://doi.org/10.1146/annurev-food-030713-092350

Begum Yasmin, A., Rydberg Hanna, A., Thorell, K., Kwak, Y.-K., Sun, L., Joffré, E., Qadri, F. and Sjöling, Å. (2018) In Situ Analyses Directly in Diarrheal Stool Reveal Large Variations in Bacterial Load and Active Toxin Expression of Enterotoxigenic *Escherichia coli* and Vibrio cholerae. *mSphere,* 3**,** https://doi.org/10.1128/msphere.00517-17. Available at: 10.1128/msphere.00517-17

Bekendam, R.H. and Flaumenhaft, R. (2019) Assays of thiol isomerase enzymatic activity. *Methods in Molecular Biology,* 1967**,** 133-148. Available at: https://doi.org/10.1007/978-1-4939-9187-7_8

Belas, A., Marques, C., Aboim, C. and Pomba, C. (2018) Emergence of *Escherichia coli* ST131 H30/H30-Rx subclones in companion animals. *Journal of Antimicrobial Chemotherapy,* 74**,** 266-269. Available at: https://doi.org/10.1093/jac/dky381

Benz, F., Huisman, J.S., Bakkeren, E., Herter, J.A., Stadler, T., Ackermann, M., Diard, M., Egli, A., Hall, A.R., Hardt, W.-D. and Bonhoeffer, S. (2021) Plasmid- and strain-specific factors drive variation in ESBL-plasmid spread *in vitro* and *in vivo*. *The ISME Journal,* 15**,** 862-878. Available at: https://doi.org/10.1038/s41396-020-00819-4

Bessonov, K., Laing, C., Robertson, J., Yong, I., Ziebell, K., Gannon, V.P.J., Nichani, A., Arya, G., Nash, J.H.E. and Christianson, S. (2021) ECTyper: in silico *Escherichia coli* serotype and species prediction from raw and assembled whole-genome sequence data. *Microbial Genomics,* 7**,** 000728. Available at: https://doi.org/10.1099/mgen.0.000728

Bielaszewska, M., Mellmann, A., Zhang, W., Köck, R., Fruth, A., Bauwens, A., Peters, G. and Karch, H. (2011) Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *The Lancet Infectious Diseases,* 11**,** 671-676. Available at: https://doi.org/10.1016/S1473-3099(11)70165-7

Biggs, M.B., Medlock, G.L., Moutinho, T.J., Lees, H.J., Swann, J.R., Kolling, G.L. and Papin, J.A. (2017) Systems-level metabolism of the altered Schaedler flora, a complete gut microbiota. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology,* 11**,** 426-438. Available at: https://doi.org/10.1038/ismej.2016.130

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., ColladoVides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science,* 277**,** 1453-1462. Available at: https://doi.org/10.1126/science.277.5331.1453

Bloomfield, S.J., Hildebrand, F., Zomer, A.L., Palau, R. and Mather, A.E. (2025) Ecological insights into the microbiology of food using metagenomics and its potential surveillance applications. *Microbial Genomics,* 11. Available at: https://doi.org/10.1099/mgen.0.001337

Bloomfield, S.J., Janecko, N., Palau, R., Alikhan, N.-F. and Mather, A.E. (2023a) Genomic diversity and epidemiological significance of non-typhoidal *Salmonella* found in retail food collected in Norfolk, UK. *Microbial Genomics,* 9**,** 001075. Available at: https://doi.org/10.1099/mgen.0.001075

Bloomfield, S.J., Palau, R., Holden, E.R., Webber, M.A. and Mather, A.E. (2024) Genomic characterization of *Pseudomonas spp*. on food: implications for spoilage, antimicrobial resistance and human infection. *BMC Microbiology,* 24**,** 20. Available at: https://doi.org/10.1186/s12866-023-03153-9

Bloomfield, S.J., Zomer, A.L., O'Grady, J., Kay, G.L., Wain, J., Janecko, N., Palau, R. and Mather, A.E. (2023b) Determination and quantification of microbial communities and antimicrobial resistance on food through host DNA-depleted metagenomics. *Food Microbiology,* 110**,** 104162. Available at: https://doi.org/10.1016/j.fm.2022.104162

Boehmer, T., Vogler, A.J., Thomas, A., Sauer, S., Hergenroether, M., Straubinger, R.K., Birdsell, D., Keim, P., Sahl, J.W., Williamson, C.H.D. and Riehm, J.M. (2018)

Phenotypic characterization and whole genome analysis of extended-spectrum beta-lactamase-producing bacteria isolated from dogs in Germany. *PLOS ONE,* 13**,** e0206252. Available at: 10.1371/journal.pone.0206252

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120. Available at: https://doi.org/10.1093/bioinformatics/btu170

Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B.B., Malhotra-Kumar, S., Westh, H., Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykasenoja, S., Olkkola, S., Wieczorek, K., Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O. and Aarestrup, F.M. (2020) ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy,* 75**,** 3491-3500. Available at: https://doi.org/10.1093/jac/dkaa345

Bosi, E., Fani, R. and Fondi, M. (2015) Defining Orthologs and Pangenome Size Metrics. *In:* Mengoni, A., Galardini, M. and Fondi, M. (ed. 1) *Bacterial Pangenomics: Methods and Protocols.* New York, NY: Springer New York.191-202. Available at: https://doi.org/10.1007/978-1-4939-1720-4_13

Bourne, J.A., Chong, W.L. and Gordon, D.M. (2019) 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,* 14**,** e0212867. Available at: https://doi.org/10.1371/journal.pone.0212867

British Meat Processors Association. (2022) *Modules & Rules 6 Pig Slaughter*. https://qualityporkcharter.org/modules-rules/. Available: https://qualityporkcharter.org/rules/pig-welfare-slaughter/pig-slaughter/ [Accessed 05/06/2024].

Brochier, C., Bapteste, E., Moreira, D. and Philippe, H. (2002) Eubacterial phylogeny based on translational apparatus proteins. *Trends in Genetics,* 18**,** 1-5. Available at: https://doi.org/10.1016/S0168-9525(01)02522-7

Brouwer, M.S.M., Zandbergen Van Essen, A., Kant, A., Rapallini, M., Harders, F., Bossers, A., Wullings, B., Wit, B. and Veldman, K.T. (2023) Implementation of WGS analysis of ESBL-producing *Escherichia coli* within EU AMR monitoring in livestock and meat. *Journal of Antimicrobial Chemotherapy***,** 1701–1704. Available at: https://doi.org/10.1093/jac/dkad158

Brower, A.V.Z. (2016) What is a cladogram and what is not? *Cladistics,* 32**,** 573-576. Available at: https://doi.org/10.1111/cla.12144

Buberg, M.L., Wasteson, Y., Lindstedt, B.A. and Witsø, I.L. (2023) *In vitro* digestion of ESC-resistant *Escherichia col*i from poultry meat and evaluation of human health risk. *Frontiers in Microbiology,* Volume 14 - 2023. Available at: https://doi.org/10.3389/fmicb.2023.1050143

Buranasinsup, S., Wiratsudakul, A., Chantong, B., Maklon, K., Suwanpakdee, S., Jiemtaweeboon, S. and Sakcamduang, W. (2023) Prevalence and characterization of antimicrobial-resistant *Escherichia coli* isolated from veterinary staff, pets, and pet owners in Thailand. *Journal of Infection and Public Health,* 16**,** 194-202. Available at: https://doi.org/10.1016/j.jiph.2023.11.006

Campitelli, E. (2022) *ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'*. Available: https://doi.org/10.5281/zenodo.2543762 [Accessed 15/04/2025].

Carattoli, A. and Hasman, H. (2020) PlasmidFinder and *in silico* pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). *In:* De La Cruz, F. (ed.) *Horizontal Gene Transfer: Methods and Protocols.* New York, NY: Springer US.285-294.

Castro-Rosas, J., Cerna-Cortes, J.F., Mendez-Reyes, E., Lopez-Hernandez, D., Gomez-Aldapa, C.A. and Estrada-Garcia, T. (2012) Presence of faecal coliforms, *Escherichia coli* and diarrheagenic *E. coli* pathotypes in ready-to-eat salads, from an area where crops are irrigated with untreated sewage water. *International Journal of Food Microbiology,* 156**,** 176-180. Available at: https://doi.org/10.1016/j.ijfoodmicro.2012.03.025

Catalano, A., Iacopetta, D., Ceramella, J., Scumaci, D., Giuzio, F., Saturnino, C., Aquaro, S., Rosano, C. and Sinicropi, M.S. (2022) Multidrug resistance (MDR): a widespread phenomenon in pharmacological therapies. *Molecules,* 27. Available at: https://doi.org/10.3390/molecules27030616

Caugant, D.A., Levin, B.R. and Selander, R.K. (1981) Genetic diversity and temporal variation in the *E. coli* population of a human host. *Genetics,* 98**,** 467-490. Available at: https://doi.org/10.1093/genetics/98.3.467

Caugant, D.A., Levin, B.R. and Selander, R.K. (1984) Distribution of multilocus genotypes of *Escherichia coli* within and between host families. *Journal of Hygiene,* 92**,** 377-84. Available at: https://doi.org/10.1017/s0022172400064597

Chan, Y., Hoban, A., Moore, H., Greig, D., Painset, A., Jorgensen, F., Chattaway, M.A., Jenkins, C., Balasegaram, S., McCormick, J. and Larkin, L. (2023) Two Outbreaks of Foodborne Gastrointestinal Infection Linked to Consumption of Imported Melons, United Kingdom, March to August 2021. *Journal of Food Protection,* 86**,** 100027. Available at: https://doi.org/10.1016/j.jfp.2022.100027

Chattaway, M.A., Schaefer, U., Tewolde, R., T.J., D. and Jenkins, C. (2017) Identification of *Escherichia coli and Shigella* Species from Whole-Genome Sequences. *Journal of Clinical Microbiology,* 55**,** 616-623. Available at: https://doi.org/10.1128/JCM.01790-16

Chen, L., Wang, L., Yassin, A.K., Zhang, J., Gong, J., Qi, K., Ganta, R.R., Zhang, Y., Yang, Y., Han, X. and Wang, C. (2018) Genetic characterization of extraintestinal *Escherichia coli* isolates from chicken, cow and swine. *AMB Express,* 8**,** 117. Available at: https://doi.org/10.1186/s13568-018-0646-8

Chen, L., Zheng, D., Liu, B., Yang, J. and Jin, Q. (2015) VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Research,* 44**,** D694-D697. Available at: 10.1093/nar/gkv1239

Chen, L.H., Zheng, D.D., Liu, B., Yang, J. and Jin, Q. (2016) VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids Res,* 44**,** D694-D697. Available at: https://doi.org/10.1093/nar/gkv1239

Cheng, L., Connor, T.R., Siren, J., Aanensen, D.M. and Corander, J. (2013) Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Molecular Biology and Evolution,* 30**,** 1224-1228. Available at: https://doi.org/10.1093/molbev/mst028

Christaki, E., Marcou, M. and Tofarides, A. (2020) Antimicrobial Resistance in Bacteria: Mechanisms, Evolution, and Persistence. *Journal of Molecular Evolution,* 88**,** 26-40. Available at: https://doi.org/10.1007/s00239-019-09914-3

Chung, H.J., Bang, W. and Drake, M.A. (2006) Stress Response of *Escherichia coli*. *Comprehensive Reviews in Food Science and Food Safety,* 5**,** 52-64. Available at: https://doi.org/10.1111/j.1541-4337.2006.00002.x

Clermont, O., Bonacorsi, S. and Bingen, E. (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Applied and Environmental Microbiology,* 66**,** 4555-4558. Available at: https://doi.org/10.1128/AEM.66.10.4555-4558.2000

Clermont, O., Christenson, J.K., Denamur, E. and Gordon, D.M. (2013) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports,* 5**,** 58-65. Available at: https://doi.org/10.1111/1758-2229.12019

Clermont, O., Condamine, B., Dion, S., Gordon, D.M. and Denamur, E. (2021) The E phylogroup of *Escherichia coli* is highly diverse and mimics the whole *E. coli* species population structure. *Environmental Microbiology,* 23**,** 7139-7151. Available at: https://doi.org/10.1111/1462-2920.15742

Clermont, O., Dixit, O.V.A., Vangchhia, B., Condamine, B., Dion, S., Bridier-Nahmias, A., Denamur, E. and Gordon, D. (2019) Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. *Environmental Microbiology,* 21**,** 3107-3117. Available at: https://doi.org/10.1111/1462-2920.14713

Clermont, O., Gordon, D. and Denamur, E. (2015) Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology,* 161**,** 980-988. Available at: https://doi.org/10.1099/mic.0.000063

Cohen, P.S., McCormick, B.A., Franklin, D.P., Burghoff, R.L. and Laux, D.C. (1991) The Role of Large Intestine Mucus in Colonization of the Mouse Large Intestine by *Escherichia coli* F-18 and *Salmonella typhimurium*. *In:* Wadström, T., Mäkelä, P. H., Svennerholm, A. M. and Wolf-Watz, H. (ed. 1) *Molecular Pathogenesis of Gastrointestinal Infections.* Boston, MA: Springer US.29-31. Available at: https://doi.org/10.1007/978-1-4684-5982-1_4

Commonwealth Scientific and Industrial Research Organisation. (2019) *Prawn stars: Five facts about Aussie prawns*. https://www.csiro.au/. Available: https://www.csiro.au/en/news/all/articles/2019/december/prawn-stars-five-facts-about-aussie-prawns [Accessed 19/06/2024].

Condamine, B., Morel-Journel, T., Tesson, F., Royer, G., Magnan, M., Bernheim, A., Denamur, E., Blanquart, F. and Clermont, O. (2025) Strain phylogroup and environmental constraints shape *Escherichia coli* dynamics and diversity over a 20-year human gut time series. *The ISME Journal,* 19**,** wrae245. Available at: https://doi.org/10.1093/ismejo/wrae245

Costa, T.R.D., Ilangovan, A., Ukleja, M., Redzej, A., Santini, J.M., Smith, T.K., Egelman, E.H. and Waksman, G. (2016) Structure of the bacterial sex F pilus reveals an assembly of a stoichiometric protein-phospholipid complex. *Cell,* 166**,** 1436-1444.e10. Available at: https://doi.org/10.1016/j.cell.2016.08.025

Coura, F.M., Diniz, S.d.A., Silva, M.X., Mussi, J.M.S., Barbosa, S.M., Lage, A.P. and Heinemann, M.B. (2015) Phylogenetic Group Determination of *Escherichia coli* Isolated from Animals Samples. *The Scientific World Journal,* 2015**,** 258424-258424. Available at: https://doi.org/10.1155/2015/258424

Cozma, A.P., Rimbu, C.M., Zendri, F., Maciuca, I.E. and Timofte, D. (2022) Clonal dissemination of extended-spectrum cephalosporin-resistant Enterobacterales between dogs and humans in households and animal shelters of Romania. *Antibiotics (Basel),* 11**,** 1242. Available at: https://doi.org/10.3390/antibiotics11091242

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J. and Harris, S.R. (2014) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research,* 43**,** e15-e15. Available at: https://doi.org/10.1093/nar/gku1196

Cunningham-Smith, P. and Emery, K. (2020) Dogs and people: exploring the human-dog connection. *Journal of Ethnobiology,* 40**,** 409-413. Available at: https://doi.org/10.2993/0278-0771-40.4.409

Dahbi, G., Mora, A., Mamani, R., López, C., Alonso, M.P., Marzoa, J., Blanco, M., Herrera, A., Viso, S., García-Garrote, F., Tchesnokova, V., Billig, M., de la Cruz, F., de Toro, M., González-López, J.J., Prats, G., Chaves, F., Martínez-Martínez, L., López-Cerezo, L., Denamur, E. and Blanco, J. (2014) Molecular epidemiology and virulence of *Escherichia coli* O16:H5-ST131: comparison with H30 and H30-Rx subclones of O25b:H4-ST131. *International Journal of Medical Microbiology,* 304**,** 1247-1257. Available at: https://doi.org/10.1016/j.ijmm.2014.10.002

Dallman, T., Cross, L., Bishop, C., Perry, N., Olesen, B., Grant, K.A. and Jenkins, C. (2013) Whole genome sequencing of an unusual serotype of Shiga toxin-producing *Escherichia coli*. *Emerging Infectious Diseases,* 19**,** 1302-4. Available at: https://doi.org/10.3201/eid1908.130016

Damborg, P., Nielsen, S.S. and Guardabassi, L. (2009) *Escherichia coli* shedding patterns in humans and dogs: insights into within-household transmission of phylotypes associated with urinary tract infections. *Epidemiology and Infection,* 137**,** 1457-1464. Available at: https://doi.org/10.1017/S095026880900226X

Daneman, N., Fridman, D., Johnstone, J., Langford, B.J., Lee, S.M., MacFadden, D.M., Mponponsuo, K., Patel, S.N., Schwartz, K.L. and Brown, K.A. (2023) Antimicrobial resistance and mortality following *E. coli* bacteremia. *EClinicalMedicine,* 56**,** 101781. Available at: https://doi.org/10.1016/j.eclinm.2022.101781

Daubin, V., Gouy, M. and Perrière, G. (2001) Bacterial molecular phylogeny using supertree approach. *Genome Informatics,* 12**,** 155-64. Available at: https://doi.org/10.11234/gi1990.12.155

Daubin, V., Moran, N.A. and Ochman, H. (2003) Phylogenetics and the cohesion of bacterial genomes. *Science,* 301**,** 829-832. Available at: https://doi.org/10.1126/science.1086568

Davis, J.A., Jackson, C.R., Fedorka-Cray, P.J., Barrett, J.B., Brousse, J.H., Gustafson, J. and Kucher, M. (2011) Anatomical distribution and genetic relatedness of antimicrobial-resistant *Escherichia coli* from healthy companion animals. *Journal of Applied Microbiology,* 110**,** 597-604. Available at: https://doi.org/10.1111/j.1365-2672.2010.04920.x

Day, M.J., Hopkins, K.L., Wareham, D.W., Toleman, M.A., Elviss, N., Randall, L., Teale, C., Cleary, P., Wiuff, C., Doumith, M., Ellington, M.J., Woodford, N. and Livermore, D.M. (2019) Extended-spectrum β-lactamase-producing *Escherichia coli* in human-derived and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. *The Lancet Infectious Diseases,* 19**,** 1325-1335. Available at: https://doi.org/10.1016/S1473-3099(19)30273-7

Decano, A.G. and Downing, T. (2019) An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. *Scientific Reports,* 9. Available at: https://doi.org/10.1038/s41598-019-54004-5

Delsuc, F., Brinkmann, H. and Philippe, H. (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics,* 6**,** 361-375. Available at: https://doi.org/10.1038/nrg1603

Deneke, C., Brendebach, H., Uelze, L., Borowiak, M., Malorny, B. and Tausch, S.H. (2021) Species-specific quality control, assembly and contamination detection in microbial isolate sequences with AQUAMIS. *Genes,* 12**,** 644. Available at: https://doi.org/10.3390/genes12050644

Department for Environment Food and Rural Affairs. (2017) *Family food 2015*. GOV.UK. Available: https://www.gov.uk/government/statistics/family-food-2015 [Accessed 25/09/2025].

Department for Environment Food and Rural Affairs. (2024) *Guidance Poultry meat marketing standards*. GOV.UK. Available: https://www.gov.uk/guidance/poultry-meat-marketing-standards [Accessed 05/06/2024].

Department for Environment Food and Rural Affairs. (2025) *Bathing Water Quality*. Available: https://environment.data.gov.uk/bwq/profiles/help-understanding-data.html [Accessed 11/06/2025].

Devane, M. (2019) *Transfer of knowledge and implementation of routine analysis of environmental samples for the presence of naturalised E. coli*. Available: https://www.envirolink.govt.nz/assets/Envirolink/Reports/1844-NLRC203.pdf [Accessed 11/08/2025].

Didelot, X. and Wilson, D.J. (2015) ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology,* 11**,** e1004041. Available at: https://doi.org/10.1371/journal.pcbi.1004041

Dinu, L.-D. and Bach, S. (2011) Induction of Viable but Nonculturable *Escherichia coli* O157:H7 in the Phyllosphere of Lettuce: a Food Safety Risk Factor. *Applied and Environmental Microbiology,* 77**,** 8295-8302. Available at: https://doi.org/10.1128/AEM.05020-11

Dixit, O.V.A., O'Brien, C.L., Pavli, P. and Gordon, D.M. (2018) Within-host evolution versus immigration as a determinant of Escherichia coli diversity in the human gastrointestinal tract. *Environmental Microbiology,* 20**,** 993-1001. Available at: https://doi.org/10.1111/1462-2920.14028

Domingues, S., da Silva, G.J. and Nielsen, K.M. (2012) Integrons: Vehicles and pathways for horizontal dissemination in bacteria. *Mobile genetic elements,* 2**,** 211-223. Available at: https://doi.org/10.4161/mge.22967

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science,* 284**,** 2124-9. Available at: https://doi.org/10.1126/science.284.5423.2124

Dupuy, B., Deghmane, A.-E. and Taha, M.-K. (2013) Chapter 63 - Type IV Prepilin Peptidase. *In:* Rawlings, Neil D. and Salvesen, Guy (ed. 1) *Handbook of Proteolytic Enzymes (Third Edition).* Academic Press.261-265. Available at: https://doi.org/10.1016/B978-0-12-382219-2.00063-6

Duriez, P., Clermont, O., Bonacorsi, S., Bingen, E., Chaventré, A., Elion, J., Picard, B. and Denamur, E. (2001) Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology,* 147**,** 1671-1676. Available at: https://doi.org/10.1099/00221287-147-6-1671

Dykhuizen, D.E. and Green, L. (1991) Recombination in *Escherichia coli* and the definition of biological species. *Journal of Bacteriology,* 173**,** 7257-7268. Available at: https://doi.org/10.1128/jb.173.22.7257-7268.1991

El Badawy, A., Hamza, D., Ahmed, Z. and Sabry, M.A. (2025) Seasonal and environmental drivers of antibiotic resistance and virulence in *Escherichia coli* from aquaculture and their public health implications. *Scientific Reports,* 15**,** 16100. Available at: https://doi.org/10.1038/s41598-025-98498-8

Elankumaran, P., Cummins Max, L., Browning Glenn, F., Marenda Marc, S., Reid Cameron, J. and Djordjevic Steven, P. (2022) Genomic and Temporal Trends in Canine ExPEC Reflect Those of Human ExPEC. *Microbiology Spectrum,* 10**,** e01291-22. Available at: https://doi.org/10.1128/spectrum.01291-22

Elankumuran, P., Browning, G.F., Marenda, M.S., Kidsley, A., Osman, M., Haenni, M., Johnson, J.R., Trott, D.J., Reid, C.J. and Djordjevic, S.P. (2023) Identification of genes influencing the evolution of Escherichia coli ST372 in dogs and humans. *Microbial Genomics,* 9. Available at: https://doi.org/10.1099/mgen.0.000930

Ellis-Iversen, J., Seyfarth, A.M., Korsgaard, H., Bortolaia, V., Munck, N. and Dalsgaard, A. (2020) Antimicrobial resistant *E. coli* and enterococci in pangasius fillets and prawns in Danish retail imported from Asia. *Food Control,* 114**,** 106958. Available at: https://doi.org/10.1016/j.foodcont.2019.106958

Escobar-Páramo, P., Giudicelli, C., Parsot, C. and Denamur, E. (2003) The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *Journal of molecular evolution,* 57**,** 140-148. Available at: https://doi.org/10.1007/s00239-003-2460-3

European Commission (2005) Commission Regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. [Accessed: 31/10/2024]

Ewers, C., Grobbel, M., Stamm, I., Kopp, P.A., Diehl, I., Semmler, T., Fruth, A., Beutlich, J., Guerra, B., Wieler, L.H. and Guenther, S. (2010) Emergence of human pandemic O25:H4-ST131 CTX-M-15 extended-spectrum-β-lactamase-producing *Escherichia coli* among companion animals. *Journal of Antimicrobial Chemotherapy,* 65**,** 651-660. Available at: https://doi.org/10.1093/jac/dkq004

Fabich, A.J., Jones, S.A., Chowdhury, F.Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J.W., Hightower, G.A., Smith, J.T., Autieri, S.M., Leatham, M.P., Lins, J.J., Allen, R.L., Laux, D.C., Cohen, P.S. and Conway, T. (2008) Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infection and immunity,* 76**,** 1143-52. Available at: https://doi.org/10.1128/IAI.01386-07

Farmery, A., Gardner, C., Green, B.S., Jennings, S. and Watson, R. (2015) Life cycle assessment of wild capture prawns: expanding sustainability considerations in the Australian Northern Prawn Fishery. *Journal of Cleaner Production,* 87**,** 96-104. Available at: https://doi.org/10.1016/j.jclepro.2014.10.063

Farooq, A., Rafique, A., Han, E. and Park, S.-M. (2025) Global dissemination of the beta-lactam resistance gene blaTEM-1 among pathogenic bacteria. *Science of The Total Environment,* 963**,** 178521. Available at: https://doi.org/10.1016/j.scitotenv.2025.178521

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution,* 17**,** 368-376. Available at: https://doi.org/10.1007/BF01734359

Food Standards Agency. (2023) *Pathogen Surveillance in Agriculture, Food and Environment Programme*. Available: https://www.food.gov.uk/our-work/pathogen-surveillance-in-agriculture-food-and-environment-programme [Accessed 06/03/2023].

Food Standards Agency. (2024) *Foodborne pathogens*. Available: https://www.food.gov.uk/research/foodborne-pathogens [Accessed 30/06/2025].

Forde, B.M., Roberts, L.W., Phan, M.-D., Peters, K.M., Fleming, B.A., Russell, C.W., Lenherr, S.M., Myers, J.B., Barker, A.P., Fisher, M.A., Chong, T.-M., Yin, W.-F., Chan, K.-G., Schembri, M.A., Mulvey, M.A. and Beatson, S.A. (2019) Population dynamics of an

Escherichia coli ST131 lineage during recurrent urinary tract infection. *Nature Communications,* 10, 3643. Available at: https://doi.org/10.1038/s41467-019-11571-5

Foster-Nyarko, E., Alikhan, N.F., Ikumapayi, U.N., Sarwar, G., Okoi, C., Tientcheu, P.M., Defernez, M., O'Grady, J., Antonio, M. and Pallen, M.J. (2021) Genomic diversity of *Escherichia coli* from healthy children in rural Gambia. *PeerJ,* 9**,** e10572. Available at: https://doi.org/10.7717/peerj.10572

Foster-Nyarko, E. and Pallen, M.J. (2022) The microbial ecology of *Escherichia coli* in the vertebrate gut. *FEMS Microbiology Reviews,* 46**,** fuac008. Available at: https://doi.org/10.1093/femsre/fuac008

Gangarosa, E.J. (1978) Epidemiology of *Escherichia coli* in the United States. *The Journal of Infectious Diseases,* 137**,** 634-638. Available at: https://doi.org/10.1093/infdis/137.5.634

Gangiredla, J., Mammel, M.K., Barnaba, T.J., Tartera, C., Gebru, S.T., Pate, I.R., Leonard, S.R., Kotewicz, M.L., Lampel, K.A., Elkins, C.A. and Lacher, D.W. (2018) Draft Genome Sequences of *Escherichia albertii, Escherichia fergusonii,* and Strains Belonging to Six Cryptic Lineages of *Escherichia* spp. *Microbiology Resource Announcements,* 6. Available at: https://doi.org/10.1128/genomeA.00271-1810.1128/genomeA.00271-18

García, A., Fox, J.G. and Besser, T.E. (2010) Zoonotic Enterohemorrhagic *Escherichia coli* : A One Health Perspective. *ILAR Journal,* 51**,** 221-232. Available at: https://doi.org/10.1093/ilar.51.3.221

Gautom, R.K. (1997) Rapid pulsed-field gel electrophoresis protocol for typing of *Escherichia coli* O157:H7 and other gram-negative organisms in 1 day. *Journal of Clinical Microbiology,* 35**,** 2977-80. Available at: https://doi.org/10.1128/jcm.35.11.2977-2980.1997

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M.V., Grechkin, Y., Mseeh, F., Fonstein, M.Y., Overbeek, R., Barabási, A.L., Oltvai, Z.N. and Osterman, A.L. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *Journal of Bacteriology,* 185**,** 5673-84. Available at: https://doi.org/10.1128/jb.185.19.5673-5684.2003

Geurtsen, J., de Been, M., Weerdenburg, E., Zomer, A., McNally, A. and Poolman, J. (2022) Genomics and pathotypes of the many faces of *Escherichia coli*. *FEMS Microbiology Reviews,* 46. Available at: https://doi.org/10.1093/femsre/fuac031

Gevers, D., Vandepoele, K., Simillion, C. and Van de Peer, Y. (2004) Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in Microbiology,* 12**,** 148-154. Available at: https://doi.org/10.1016/j.tim.2004.02.007

Giufre, M., Accogli, M., Farina, C., Giammanco, A., Pecile, P. and Cerquetti, M. (2014) Predominance of the fimH30 Subclone Among Multidrug-Resistant *Escherichia coli* Strains Belonging to Sequence Type 131 in Italy. *The Journal of Infectious Diseases,* 209**,** 629-630. Available at: https://doi.org/10.1093/infdis/jit583

Goldman, A.D. and Landweber, L.F. (2016) What Is a Genome? *PLoS genetics,* 12**,** e1006181-e1006181. Available at: https://doi.org/10.1371/journal.pgen.1006181

Gonzalez-Alba, J.M., Baquero, F., Cantón, R. and Galán, J.C. (2019) Stratified reconstruction of ancestral *Escherichia coli* diversification. *BMC Genomics,* 20**,** 936. Available at: https://doi.org/10.1186/s12864-019-6346-1

Gordo, I., Demengeot, J. and Xavier, K. (2014) Escherichia coli adaptation to the gut environment: a constant fight for survival. *Future Microbiology,* 9**,** 1235-1238. Available at: https://doi.org/10.2217/fmb.14.86

GOV UK. (2009) *Treating and Processing Poultry*. Environment Agency. Available: https://www.gov.uk/government/publications/treating-and-processing-poultry-additional-guidance [Accessed 11/08/2025].

GOV UK 2021. United Kingdom Food Security Report 2021: Theme 5: Food Safety and Consumer Confidence. London, UK.

GOV UK. (2024) *Confronting Antimicrobial Resistance: 2024 to 2029*. Department of Health and Social Care. Available: https://www.gov.uk/government/publications/uk-5-year-action-plan-for-antimicrobial-resistance-2024-to-2029/confronting-antimicrobial-resistance-2024-to-2029 [Accessed 11/06/2025].

Groat, E.F., Williams, N.J., Pinchbeck, G., Warner, B., Simpson, A. and Schmidt, V.M. (2022) UK dogs eating raw meat diets have higher risk of Salmonella and antimicrobial-resistant *Escherichia col*i faecal carriage. *Journal of small animal practice,* 63**,** 435-441. Available at: https://doi.org/10.1111/jsap.13488

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-1075. Available at: https://doi.org/10.1093/bioinformatics/btt086

Guttman, D.S. and Dykhuizen, D.E. (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science,* 266**,** 1380-3. Available at: https://doi.org/10.1126/science.7973728

Guy, R.L., Rudman, J., Higgins, H., Carter, E., Henderson, K.L., Demirjian, A. and Gerver, S.M. (2023) Nitrofurantoin resistance as an indicator for multidrug resistance: an assessment of *Escherichia coli* urinary tract specimens in England, 2015–19. *JAC-Antimicrobial Resistance,* 5. Available at: https://doi.org/10.1093/jacamr/dlad122

Halkman, H.B.D. and Halkman, A.K. (2014) Indicator Organisms. *In:* Batt, Carl A. and Tortorello, Mary Lou (ed. 2) *Encyclopedia of Food Microbiology (Second Edition).* Oxford: Academic Press.358-363. Available at: https://doi.org/10.1016/B978-0-12-384730-0.00396-7

Hall, R.J., Whelan, F.J., Cummins, E.A., Connor, C., McNally, A. and McInerney, J.O. (2021) Gene-gene relationships in an *Escherichia coli* accessory genome are linked to function and mobility. *Microbial Genomics,* 7. Available at: https://doi.org/10.1099/mgen.0.000650

Hamilton-Miller, J.M. (2005) Correct use of the word 'fomites': its etymology is important. *The Journal of hospital infection,* 59**,** 165. Available at: https://doi.org/10.1016/j.jhin.2004.08.004

Hansen, W. and Yourassowsky, E. (1984) Detection of beta-glucuronidase in lactose-fermenting members of the family Enterobacteriaceae and its presence in bacterial urine cultures. *Journal of Clinical Microbiology,* 20**,** 1177-1179. Available at: https://doi.org/10.1128/jcm.20.6.1177-1179.1984

Hara-Kudo, Y. and Takatori, K. (2011) Contamination level and ingestion dose of foodborne pathogens associated with infections. *Epidemiology and Infection,* 139**,** 1505-1510. Available at: https://doi.org/10.1017/S095026881000292X

Harada, K., Okada, E., Shimizu, T., Kataoka, Y., Sawada, T. and Takahashi, T. (2012) Antimicrobial resistance, virulence profiles, and phylogenetic groups of fecal *Escherichia coli* isolates: a comparative analysis between dogs and their owners in Japan. *Comparative Immunology, Microbiology and Infectious Diseases,* 35**,** 139-144. Available at: https://doi.org/10.1016/j.cimid.2011.12.005

Hasegawa, M., Kishino, H. and Saitou, N. (1991) On the maximum likelihood method in molecular phylogenetics. *Journal of Molecular Evolution,* 32**,** 443-445. Available at: https://doi.org/10.1007/BF02101285

Hayashi, W., Ohsaki, Y., Taniguchi, Y., Koide, S., Kawamura, K., Suzuki, M., Kimura, K., Wachino, J.I., Nagano, Y., Arakawa, Y. and Nagano, N. (2018) High prevalence of bla(CTX-M-14) among genetically diverse *Escherichia coli* recovered from retail raw chicken meat portions in Japan. *International Journal of Food Microbiology,* 284**,** 98-104. Available at: https://doi.org/10.1016/j.ijfoodmicro.2018.08.003

Hayer, J., Salgado-Caxito, M., Opazo-Capurro, A., Muñoz, P.G., Millán, J., Piñeiro, A., Munita, J.M., Rivas, L. and Benavides, J.A. (2023) Multiple clonal transmissions of clinically relevant extended-spectrum beta-lactamase–producing *Escherichia coli* among livestock, dogs, and wildlife in Chile. *Journal of Global Antimicrobial Resistance,* 34**,** 247-252. Available at: https://doi.org/10.1016/j.jgar.2023.07.009

Hazen, T.H., Michalski, J.M., Tennant, S.M. and Rasko, D.A. (2023) Genomic diversity of non-diarrheagenic fecal *Escherichia coli* from children in sub-Saharan Africa and south Asia and their relatedness to diarrheagenic *E. coli*. *Nature Communications,* 14**,** 1400. Available at: https://doi.org/10.1038/s41467-023-36337-y

Health Protection Agency. (2009) *Guidelines for assessing the microbiological safety of ready-to-eat foods*. Health Protection Agency. Available: https://www.gov.uk/government/publications/ready-to-eat-foods-microbiological-safety-assessment-guidelines [Accessed 31/10/2024].

Heather, J.M. and Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8. Available at: https://doi.org/10.1016/j.ygeno.2015.11.003

Heiden, S.E., Kurz, M.S.E., Bohnert, J., Bayingana, C., Ndoli, J.M., Sendegeya, A., Gahutu, J.B., Eger, E., Mockenhaupt, F.P. and Schaufler, K. (2020) Flies from a tertiary hospital in Rwanda carry multidrug-resistant Gram-negative pathogens including extended-spectrum beta-lactamase-producing *E. coli* sequence type 131. *Antimicrob Resist Infect Control,* 9**,** 34. Available at: https://doi.org/10.1186/s13756-020-0696-y

Holland, D. and Mahmoudzadeh, N. (2020) *Foodborne disease estimates for the United Kingdom in 2018*. Food Standards Agency. Available: https://www.food.gov.uk/research/foodborne-disease/foodborne-disease-estimates-for-the-united-kingdom-in-2018 [Accessed 11/08/2025].

Holland, R.E., Walker, R.D., Sriranganathan, N., Wilson, R.A. and Ruhl, D.C. (1999) Characterization of *Escherichia coli* isolated from healthy dogs. *Veterinary Microbiology,* 70**,** 261-268. Available at: https://doi.org/10.1016/S0378-1135(99)00136-4

Holley, G. and Melsted, P. (2019) Bifrost – Highly parallel construction and indexing of colored and compacted de Bruijn graphs. *bioRxiv***,** 695338. Available at: https://doi.org/10.1186/s13059-020-02135-8

Holvoet, K., De Keuckelaere, A., Sampers, I., Van Haute, S., Stals, A. and Uyttendaele, M. (2014) Quantitative study of cross-contamination with *Escherichia col*i, *E. coli* O157, MS2 phage and murine norovirus in a simulated fresh-cut lettuce wash process. *Food Control,* 37**,** 218-227. Available at: https://doi.org/10.1016/j.foodcont.2013.09.051

Holvoet, K., Jacxsens, L., Sampers, I. and Uyttendaele, M. (2012) Insight into the Prevalence and Distribution of Microbial Contamination To Evaluate Water Management in the Fresh Produce Processing Industry. *Journal of Food Protection,* 75**,** 671-681. Available at: https://doi.org/10.4315/0362-028X.JFP-11-175

Hsieh, T., Ma, K. and Chao, A. (2016) iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution,* 7**,** 1451-1456. Available at: https://doi.org/10.1111/2041-210X.12613

Hunt, M., Mather, A.E., Sanchez-Buso, L., Page, A.J., Parkhill, J., Keane, J.A. and Harris, S.R. (2017) ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial Genomics,* 3**,** e000131. Available at: https://doi.org/10.1099/mgen.0.000131

Hussain, A., Ranjan, A., Nandanwar, N., Babbar, A., Jadhav, S. and Ahmed, N. (2014) Genotypic and phenotypic profiles of *Escherichia coli* isolates belonging to clinical sequence type 131 (ST131), clinical non-ST131, and fecal non-ST131 lineages from India. *Antimicrobial agents and chemotherapy,* 58**,** 7240-9. Available at: https://doi.org/10.1128/AAC.03320-14

International Organization for Standardization (2001) ISO 16649-2:2001. Microbiology of food and animal feeding stuffs — Horizontal method for the enumeration of beta-glucuronidase-positive *Escherichia coli* - Part 2: Colony-count technique at 44 degrees C using 5-bromo-4-chloro-3-indolyl beta-D-glucuronide. ISO 16649-2:2001.

International Organization for Standardization (2015) ISO 16649-3:2015. Microbiology of the food chain — Horizontal method for the enumeration of beta-glucuronidase-positive *Escherichia coli* - Part 3: Detection and most probable number technique using 5-bromo-4-chloro-3-indolyl-ß-D-glucuronide. ISO 16649-3:2015.

Ishii, S., Ksoll, W.B., Hicks, R.E. and Sadowsky, M.J. (2006) Presence and growth of naturalized *Escherichia coli* in temperate soils from lake superior watersheds. *Applied and Environmental Microbiology,* 72**,** 612-621. Available at: https://doi.org/10.1128/AEM.72.1.612-621.2006

Islam, M., Doyle, M.P., Phatak, S.C., Millner, P. and Jiang, X. (2004) Persistence of enterohemorrhagic *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water. *Journal of Food Protection,* 67**,** 1365-1370. Available at: https://doi.org/10.4315/0362-028X-67.7.1365

Jackson, R.W., Vinatzer, B., Arnold, D.L., Dorus, S. and Murillo, J. (2011) The influence of the accessory genome on bacterial pathogen evolution. *Mobile genetic elements,* 1**,** 55-65. Available at: https://doi.org/10.4161/mge.1.1.16432

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T. and Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications,* 9. Available at: https://doi.org/10.1038/s41467-018-07641-9

Janecko, N., Zamudio, R., Palau, R., Bloomfield, S.J. and Mather, A.E. (2023) Repeated cross-sectional study identifies differing risk factors associated with microbial contamination in common food products in the United Kingdom. *Food Microbiology,* 111**,** 104196. Available at: https://doi.org/10.1016/j.fm.2022.104196

Jasper, J., Elmore, J.S. and Wagstaff, C. (2021) Determining the quality of leafy salads: Past, present and future. *Postharvest Biology and Technology,* 180**,** 111630. Available at: https://doi.org/10.1016/j.postharvbio.2021.111630

Jensen, D.A., Friedrich, L.M., Harris, L.J., Danyluk, M.D. and Schaffner, D.W. (2015) Cross contamination of Escherichia coli O157:H7 between lettuce and wash water during home-scale washing. *Food Microbiology,* 46**,** 428-433. Available at: https://doi.org/10.1016/j.fm.2014.08.025

Jesser, K.J., Trueba, G., Konstantinidis, K.T. and Levy, K. (2023) Why are so many enteric pathogen infections asymptomatic? Pathogen and gut microbiome characteristics associated with diarrhea symptoms and carriage of diarrheagenic *E. coli* in northern Ecuador. *Gut Microbes,* 15**,** 2281010. Available at: https://doi.org/10.1080/19490976.2023.2281010

Johnson, J., Johnston, B., Porter, S., Clabots, C., Bender, T., Thuras, P., Trott, D., Cobbold, R., Mollinger, J., Ferrieri, P., Drawz, S. and Banerjee, R. (2019) Rapid emergence, subsidence, and molecular detection of *Escherichia coli* sequence type 1193-*fimH*64, a new disseminated multidrug-resistant commensal and extraintestinal pathogen. *Journal of Clinical Microbiology,* 57**,** e01664-18. Available at: https://doi.org/10.1128/jcm.01664-18

Johnson, J.R. and Clabots, C. (2006) Sharing of virulent *Escherichia coli* clones among household members of a woman with acute cystitis. *Clinical Infectious Diseases,* 43**,** e101-e108. Available at: https://doi.org/10.1086/508541

Johnson, J.R., Clabots, C. and Kuskowski, M.A. (2008) Multiple-host sharing, long-term persistence, and virulence of *Escherichia coli* clones from human and animal household members. *Journal of Clinical Microbiology,* 46**,** 4078-4082. Available at: https://doi.org/10.1128/JCM.00980-08

Johnson, J.R., Davis, G., Clabots, C., Johnston, B.D., Porter, S., DebRoy, C., Pomputius, W., Ender, P.T., Cooperstock, M., Slater, B.S., Banerjee, R., Miller, S., Kisiela, D., Sokurenko, E.V., Aziz, M. and Price, L.B. (2016) Household clustering of *Escherichia coli* sequence type 131 clinical and fecal isolates according to whole genome sequence analysis. *Open Forum Infectious Diseases,* 3**,** ofw129. Available at: https://doi.org/10.1093/ofid/ofw129

Johnson, J.R., Delavari, P., Stell, A.L., Whittam, T.S., Carlino, U. and Russo, T.A. (2001) Molecular Comparison of Extraintestinal *Escherichia coli* Isolates of the Same Electrophoretic Lineages from Humans and Domestic Animals. *The Journal of Infectious Diseases,* 183**,** 154-159. Available at: https://doi.org/10.1086/317662

Johnson, J.R., Murray, A.C., Gajewski, A., Sullivan, M., Snippes, P., Kuskowski, M.A. and Smith, K.E. (2003) Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic *Escherichia coli* from retail chicken products. *Antimicrobial Agents and Chemotherapy,* 47**,** 2161-2168. Available at: https://doi.org/10.1128/AAC.47.7.2161-2168.2003

Johnson Timothy, J., Wannemuehler Yvonne, M., Johnson Sara, J., Logue Catherine, M., White David, G., Doetkott, C. and Nolan Lisa, K. (2007) Plasmid Replicon Typing of Commensal and Pathogenic *Escherichia coli* Isolates. *Applied and Environmental Microbiology,* 73**,** 1976-1983. Available at: https://doi.org/10.1128/AEM.02171-06

Johnson, T.J., Armstrong, J.R., Johnston, B., Merino-Velasco, I., Jamborova, I., Singer, R.S., Johnson, J.R. and Bender, J.B. (2022) Occurrence and potential transmission of extended-spectrum beta-lactamase-producing extraintestinal pathogenic and enteropathogenic *Escherichia coli* in domestic dog faeces from Minnesota. *Zoonoses and Public Health,* 69**,** 888-895. Available at: https://doi.org/10.1111/zph.12985

Juraschek, K., Borowiak, M., Tausch, S.H., Malorny, B., Käsbohrer, A., Otani, S., Schwarz, S., Meemken, D., Deneke, C. and Hammerl, J.A. (2021) Outcome of Different Sequencing and Assembly Approaches on the Detection of Plasmids and Localization of Antimicrobial Resistance Genes in Commensal *Escherichia coli*. *Microorganisms,* 9. Available at: https://doi.org/10.3390/microorganisms9030598

Kaas, R.S., Friis, C., Ussery, D.W. and Aarestrup, F.M. (2012) Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics,* 13. Available at: https://doi.org/10.1186/1471-2164-13-577

Karahutová, L., Mandelík, R. and Bujňáková, D. (2021) Antibiotic Resistant and Biofilm-Associated *Escherichia coli* Isolates from Diarrheic and Healthy Dogs. *Microorganisms,* 9. Available at: https://doi.org/10.3390/microorganisms9061334

Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovacs, K., Mehi, O., Baliko, G., Szappanos, B., Gyorfy, Z., Feher, T., Bogos, B., Blattner, F.R., Pal, C., Posfai, G. and Papp, B. (2016) Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Molecular Biology and Evolution,* 33**,** 1257-1269. Available at: https://doi.org/10.1093/molbev/msw009

Katz, L.S., Griswold, T., Morrison, S.S., Caravas, J.A., Zhang, S., den Bakker, H.C., Deng, X. and Carleton, H.A. (2019) Mashtree: a rapid comparison of whole genome sequence files. *Journal of Open Source Software,* 4. Available at: https://doi.org/10.21105/joss.01762

Kauffmann, F. (1947) The Serology of the Coli Group. *The Journal of Immunology,* 57**,** 71-100. Available at: https://doi.org/10.4049/jimmunol.57.1.71

Kaushik, A., Taneja, N.K., Chauhan, K. and Oberoi, H.S. (2025) Dynamics of mono and dual-species biofilms of *Escherichia coli* and *Salmonella Typhimurium*: Interspatial interactions and novel inhibition strategies. *International Journal of Food Microbiology,* 440**,** 111280. Available at: https://doi.org/10.1016/j.ijfoodmicro.2025.111280

Kay, D. and Fricker, C. (1997) Coliforms and *E. coli* : problem or solution? *In:* Kay, David and Fricker, Colin (ed. 1) *Chapter 11 - Microbiological Indicators of Recreational Water Quality.* The Royal Society of Chemistry: Available at: https://research.ebsco.com/linkprocessor/plink?id=3d95b78a-5eb9-3bfb-a85e-78b351460cb9

Keenan, D.F. (2016) Pork Meat Quality, Production and Processing on. *In:* Caballero, Benjamin, Finglas, Paul M. and Toldrá, Fidel (ed. *Encyclopedia of Food and Health.* Oxford: Academic Press.419-431. Available at: https://doi.org/10.1016/B978-0-12-384947-2.00551-1

Kidsley, A.K., O'Dea, M., Saputra, S., Jordan, D., Johnson, J.R., Gordon, D.M., Turni, C., Djordjevic, S.P., Abraham, S. and Trott, D.J. (2020a) Genomic analysis of phylogenetic group B2 extraintestinal pathogenic *E. coli* causing infections in dogs in Australia. *Veterinary Microbiology,* 248**,** 108783. Available at: https://doi.org/10.1016/j.vetmic.2020.108783

Kidsley, A.K., White, R.T., Beatson, S.A., Saputra, S., Schembri, M.A., Gordon, D., Johnson, J.R., O'Dea, M., Mollinger, J.L., Abraham, S. and Trott, D.J. (2020b) Companion animals are spillover hosts of the multidrug-resistant human extraintestinal *Escherichia coli* pandemic clones ST131 and ST1193. *Frontiers in Microbiology,* 11**,** 1968. Available at: https://doi.org/10.3389/fmicb.2020.01968

Kim, C., Pongpanich, M. and Porntaveetus, T. (2024) Unraveling metagenomics through long-read sequencing: a comprehensive review. *Journal of Translational Medicine,* 22**,** 111. Available at: https://doi.org/10.1186/s12967-024-04917-1

Kim, D., Song, L., Breitwieser, F.P. and Salzberg, S.L. (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research,* 26**,** 1721-1729. Available at: https://doi.org/10.1101/gr.210641.116

Kislichkina, A.A., Kartsev, N.N., Skryabin, Y.P., Sizova, A.A., Kanashenko, M.E., Teymurazov, M.G., Kuzina, E.S., Bogun, A.G., Fursova, N.K., Svetoch, E.A. and Dyatlov, I.A. (2022) Genomic analysis of a hybrid enteroaggregative hemorrhagic *Escherichia coli* O181:H4 strain causing colitis with hemolytic-uremic syndrome. *Antibiotics (Basel),* 11. Available at: https://doi.org/10.3390/antibiotics11101416

Klotzl, F. and Haubold, B. (2020) Phylonium: fast estimation of evolutionary distances from large samples of similar genomes. *Bioinformatics,* 36**,** 2040-2046. Available at: https://doi.org/10.1093/bioinformatics/btz903

Korajkic, A. and Harwood, V. (2016) Water supplies: microbiological analysis. Available at: https://doi.org/10.1016/B978-0-12-384947-2.00741-8

Kothary, M.H. and Babu, U.S. (2001) Infective dose of foodborne pathogens in volunteers: A review. *Journal of Food Safety,* 21**,** 49-68. Available at: https://doi.org/10.1111/j.1745-4565.2001.tb00307.x

Krause, G., Zimmermann, S. and Beutin, L. (2005) Investigation of domestic animals and pets as a reservoir for intimin- (eae) gene positive *Escherichia coli* types. *Veterinary Microbiology,* 106**,** 87-95. Available at: https://doi.org/10.1016/j.vetmic.2004.11.012

Kuhner, M.K. and Felsenstein, J. (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution,* 11**,** 459-468. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a040126

Lambrecht, E., Van Coillie, E., Van Meervenne, E., Boon, N., Heyndrickx, M. and Van de Wiele, T. (2019) Commensal *E. coli* rapidly transfer antibiotic resistance genes to human intestinal microbiota in the Mucosal Simulator of the Human Intestinal Microbial Ecosystem (M-SHIME). *International Journal of Food Microbiology,* 311**,** 108357. Available at: https://doi.org/10.1016/j.ijfoodmicro.2019.108357

Lang, J.M., Darling, A.E. and Eisen, J.A. (2013) Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. *PloS one,* 8**,** e62510. Available at: https://doi.org/10.1371/journal.pone.0062510

Lapierre, P., Lasek-Nesselquist, E. and Gogarten, J.P. (2014) The impact of HGT on phylogenomic reconstruction methods. *Briefings in Bioinformatics,* 15**,** 79-90. Available at: https://doi.org/10.1093/bib/bbs050

Lawrence, J.G. and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences,* 95**,** 9413-9417. Available at: https://doi.org/10.1073/pnas.95.16.9413

Lawrence, J.G. and Retchless, A.C. (2009) The Interplay of Homologous Recombination and Horizontal Gene Transfer in Bacterial Speciation. *In:* Gogarten, Maria Boekels, Gogarten, Johann Peter and Olendzenski, Lorraine C. (ed. *Horizontal Gene Transfer: Genomes in Flux*. Totowa, NJ: Humana Press.29-53. Available at: https://doi.org/10.1007/978-1-60327-853-9_3

Leatham, M.P., Stevenson, S.J., Gauger, E.J., Krogfelt, K.A., Lins, J.J., Haddock, T.L., Autieri, S.M., Conway, T. and Cohen, P.S. (2005) Mouse Intestine Selects Nonmotile flhDC Mutants of *Escherichia coli* MG1655 with Increased Colonizing Ability and Better Utilization of Carbon Sources. *Infection and Immunity,* 73**,** 8039-8049. Available at: https://doi.org/10.1128/IAI.73.12.8039-8049.2005

Lebret, B. and Čandek-Potokar, M. (2022) Review: Pork quality attributes from farm to fork. Part II. Processed pork products. *Animal,* 16**,** 100383. Available at: https://doi.org/10.1016/j.animal.2021.100383

Lee, M.S.Y. and Palci, A. (2015) Morphological Phylogenetics in the Genomic Age. *Current Biology,* 25**,** R922-R929. Available at: https://doi.org/10.1016/j.cub.2015.07.009

Lees, J.A., Galardini, M., Bentley, S.D., Weiser, J.N. and Corander, J. (2018a) pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics,* 34**,** 4310-4312. Available at: https://doi.org/10.1093/bioinformatics/bty539

Lees, J.A., Kendall, M., Parkhill, J., Colijn, C., Bentley, S.D. and Harris, S.R. (2018b) Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation based study. *Wellcome open research,* 3**,** 33-33. Available at: https://doi.org/10.12688/wellcomeopenres.14265.2

Leimbach, A., Hacker, J. and Dobrindt, U. (2013) *E. coli* as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity. *Between Pathogenicity and Commensalism,* 358**,** 3-32. Available at: https://doi.org/10.1007/82_2012_303

Leite-Martins, L., Meireles, D., Beça, N., Bessa, L.J., de Matos, A.J.F. and Martins da Costa, P. (2015) Spread of multidrug-resistant *Escherichia coli* within domestic aggregates (humans, pets, and household environment). *Journal of Veterinary Behavior,* 10**,** 549-555. Available at: https://doi.org/10.1016/j.jveb.2015.07.040

Lemoine, F. and Gascuel, O. (2021) Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genomics and Bioinformatics,* 3. Available at: https://doi.org/10.1093/nargab/lqab083

Lescat, M., Clermont, O., Woerther, P.L., Glodt, J., Dion, S., Skurnik, D., Djossou, F., Dupont, C., Perroz, G., Picard, B., Catzeflis, F., Andremont, A. and Denamur, E. (2013) Commensal *Escherichia coli* strains in Guiana reveal a high genetic diversity with host-dependant population structure. *Environmental Microbiology Reports,* 5**,** 49-57. Available at: https://doi.org/10.1111/j.1758-2229.2012.00374.x

Leschen, W., Little, D., Bunting, S. and van Veenhuizen, R. (2005) *Urban aquatic production*. Available: https://assets.publishing.service.gov.uk/media/57a09def40f0b64974001b30/R8365Pap1.pdf [Accessed 11/08/2025].

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-1760. Available at: https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079. Available at: https://doi.org/10.1093/bioinformatics/btp352

Li, X.-P., Sun, R.-Y., Song, J.-Q., Fang, L.-X., Zhang, R.-M., Lian, X.-L., Liao, X.-P., Liu, Y.-H., Lin, J. and Sun, J. (2020) Within-host heterogeneity and flexibility of mcr-1 transmission in chicken gut. *International Journal of Antimicrobial Agents,* 55**,** 105806. Available at: https://doi.org/10.1016/j.ijantimicag.2019.09.010

Liu, B., Furevi, A., Perepelov, A.V., Guo, X., Cao, H., Wang, Q., Reeves, P.R., Knirel, Y.A., Wang, L. and Widmalm, G. (2020) Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiology Reviews,* 44**,** 655-683. Available at: https://doi.org/10.1093/femsre/fuz028

Liu, B., Zheng, D., Zhou, S., Chen, L. and Yang, J. (2021) VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Research,* 50**,** D912-D917. Available at: https://doi.org/10.1093/nar/gkab1107

Liu, S., Jin, D., Lan, R., Wang, Y., Meng, Q., Dai, H., Lu, S., Hu, S. and Xu, J. (2015) *Escherichia marmotae* sp. nov., isolated from faeces of Marmota himalayana. *International Journal of Systematic and Evolutionary Microbiology,* 65**,** 2130-2134. Available at: https://doi.org/10.1099/ijs.0.000228

Liu, Y.-Y., Lu, L., Yue, C., Gao, X., Chen, J., Gao, G., Li, K., Deng, H. and Liu, J.-H. (2024) Emergence of plasmid-mediated high-level tigecycline resistance gene tet(X4) in Enterobacterales from retail aquatic products. *Food Research International,* 178**,** 113952. Available at: https://doi.org/10.1016/j.foodres.2024.113952

López-Vergé, S., Gasa, J., Farré, M., Coma, J., Bonet, J. and Solà-Oriol, D. (2018) Potential risk factors related to pig body weight variability from birth to slaughter in commercial conditions. *Translational Animal Science,* 2**,** 383-395. Available at: https://doi.org/10.1093/tas/txy082

Lu, Y., Li, M., Gao, Z., Ma, H., Chong, Y., Hong, J., Wu, J., Wu, D., Xi, D. and Deng, W. (2025) Advances in Whole Genome Sequencing: Methods, Tools, and Applications in Population Genomics. *International journal of molecular sciences,* 26. Available at: https://doi.org/10.3390/ijms26010372

Ludden, C., Coll, F., Gouliouris, T., Restif, O., Blane, B., Blackwell, G.A., Kumar, N., Naydenova, P., Crawley, C., Brown, N.M., Parkhill, J. and Peacock, S.J. (2021) Defining nosocomial transmission of *Escherichia coli* and antimicrobial resistance genes: a genomic surveillance study. *The Lancet Microbe,* 2**,** e472-e480. Available at: https://doi.org/10.1016/S2666-5247(21)00117-8

Ludden, C., Raven, K.E., Jamrozy, D., Gouliouris, T., Blane, B., Coll, F., de Goffau, M., Naydenova, P., Horner, C., Hernandez-Garcia, J., Wood, P., Hadjirin, N., Radakovic, M., Brown, N.M., Holmes, M., Parkhill, J. and Peacock, S.J. (2019) One Health Genomic Surveillance of *Escherichia coli* Demonstrates Distinct Lineages and Mobile Genetic Elements in Isolates from Humans versus Livestock. *Mbio,* 10. Available at: https://doi.org/10.1128/mbio.02693-18

Lukjancenko, O., Wassenaar, T.M. and Ussery, D.W. (2010) Comparison of 61 Sequenced *Escherichia coli* Genomes. *Microbial Ecology,* 60**,** 708-720. Available at: https://doi.org/10.1007/s00248-010-9717-3

Luo, Y., He, Q. and McEvoy, J.L. (2010) Effect of Storage Temperature and Duration on the Behavior of *Escherichia coli* O157:H7 on Packaged Fresh-Cut Salad Containing Romaine and Iceberg Lettuce. *Journal of Food Science,* 75**,** M390-M397. Available at: https://doi.org/10.1111/j.1750-3841.2010.01722.x

Magiorakos, A.P., Srinivasan, A., Carey, R.B., Carmeli, Y., Falagas, M.E., Giske, C.G., Harbarth, S., Hindler, J.F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D.L., Rice, L.B., Stelling, J., Struelens, M.J., Vatopoulos, A., Weber, J.T. and Monnet, D.L. (2012) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection,* 18**,** 268-81. Available at: https://doi.org/10.1111/j.1469-0691.2011.03570.x

Maguire, M., Kase, J.A., Roberson, D., Muruvanda, T., Brown, E.W., Allard, M., Musser, S.M. and González-Escalona, N. (2021) Precision long-read metagenomics sequencing for food safety by detection and assembly of Shiga toxin-producing *Escherichia coli* in irrigation water. *PLoS One,* 16**,** e0245172. Available at: 10.1371/journal.pone.0245172

Mahazu, S., Prah, I., Ayibieke, A., Sato, W., Hayashi, T., Suzuki, T., Iwanaga, S., Ablordey, A. and Saito, R. (2021) Possible Dissemination of *Escherichia coli* Sequence Type 410 Closely Related to B4/H24RxC in Ghana. *Frontiers in Microbiology,* 12**,** 770130. Available at: https://doi.org/10.3389/fmicb.2021.770130

Manges, A.R. and Johnson, J.R. (2012) Food-Borne Origins of Escherichia coli Causing Extraintestinal Infections. *Clinical Infectious Diseases,* 55**,** 712-719. Available at: https://doi.org/10.1093/cid/cis502

Manges, A.R., Johnson, J.R., Foxman, B., O'Bryan, T.T., Fullerton, K.E. and Riley, L.W. (2001) Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. *The New England journal of medicine,* 345**,** 1007-13. Available at: https://doi.org/10.1056/NEJMoa011265

Martin, I.E., Tyler, S.D., Tyler, K.D., Khakhria, R. and Johnson, W.M. (1996) Evaluation of ribotyping as epidemiologic tool for typing *Escherichia coli* serogroup O157 isolates. *Journal of clinical microbiology,* 34**,** 720-3. Available at: https://doi.org/10.1128/jcm.34.3.720-723.1996

Matsumura, Y., Pitout, J.D.D., Peirano, G., DeVinney, R., Noguchi, T., Yamamoto, M., Gomi, R., Matsuda, T., Nakano, S., Nagao, M., Tanaka, M. and Ichiyama, S. (2017) Rapid identification of different *Escherichia coli* sequence type 131 clades. *Antimicrobial Agents and Chemotherapy,* 61**,** e00179-17. Available at: https://doi.org/10.1128/aac.00179-17

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J., Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T. and Wright, G.D. (2013) The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy,* 57**,** 3348-57. Available at: https://doi.org/10.1128/AAC.00419-13

McInerney, J.O., McNally, A. and O'Connell, M.J. (2017) Why prokaryotes have pangenomes. *Nature Microbiology,* 2. Available at: https://doi.org/10.1038/nmicrobiol.2017.40

McKinnon, J., Roy Chowdhury, P. and Djordjevic, S.P. (2018) Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis. *International Journal of*

Antimicrobial Agents, 52**,** 430-435. Available at: https://doi.org/10.1016/j.ijantimicag.2018.06.017

McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., Vehkala, M., Valimaki, N., Prentice, M.B., Ashour, A., Avram, O., Pupko, T., Dobrindt, U., Literak, I., Guenther, S., Schaufler, K., Wieler, L.H., Zong, Z.Y., Sheppard, S.K., McInerney, J.O. and Corander, J. (2016) Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. *Plos Genetics,* 12. Available at: https://doi.org/10.1371/journal.pgen.1006280

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. (2005) The microbial pan-genome. *Current Opinion in Genetics & Development,* 15**,** 589-594. Available at: https://doi.org/10.1016/j.gde.2005.09.006

Medlock, G.L., Carey, M.A., McDuffie, D.G., Mundy, M.B., Giallourou, N., Swann, J.R., Kolling, G.L. and Papin, J.A. (2018) Inferring metabolic mechanisms of interaction within a defined gut microbiota. *Cell systems,* 7**,** 245-257. e7. Available at: https://doi.org/10.1016/j.cels.2018.08.003

Menezes, J., Frosini, S.-M., Belas, A., Marques, C., da Silva, J.M., Amaral, A.J., Loeffler, A. and Pomba, C. (2023) Longitudinal study of ESBL/AmpC-producing Enterobacterales strains sharing between cohabiting healthy companion animals and humans in Portugal and in the United Kingdom. *European Journal of Clinical Microbiology & Infectious Diseases,* 42**,** 1011-1024. Available at: https://doi.org/10.1007/s10096-023-04629-2

Menezes, J., Moreira da Silva, J., Frosini, S.-M., Loeffler, A., Weese, S., Perreten, V., Schwarz, S., Telo da Gama, L., Amaral, A.J. and Pomba, C. (2022) mcr-1 colistin resistance gene sharing between *Escherichia coli* from cohabiting dogs and humans, Lisbon, Portugal, 2018 to 2020. *Eurosurveillance,* 27**,** 2101144. Available at: https://doi.org/10.2807/1560-7917.ES.2022.27.44.2101144

Mikhail, A.F.W., Jenkins, C., Dallman, T.J., Inns, T., Douglas, A., Martín, A.I.C., Fox, A., Cleary, P., Elson, R. and Hawker, J. (2018) An outbreak of Shiga toxin-producing *Escherichia coli* O157:H7 associated with contaminated salad leaves: epidemiological, genomic and food trace back investigations. *Epidemiology and Infection,* 146**,** 187-196. Available at: https://doi.org/10.1017/S0950268817002874

Mohapatra, S.S., Dwibedy, S.K. and Padhy, I. (2021) Polymyxins, the last-resort antibiotics: Mode of action, resistance emergence, and potential solutions. *Journal of Biosciences,* 46**,** 85. Available at: https://doi.org/10.1007/s12038-021-00209-8

Moinet, M., Gadd, L., Devane, M., Wood, D., Gilpin, B. and Cookson, A. (2021) *Faecal source tracking and the identification of naturalised Escherichia coli to assist with* *establishing water quality and faecal contamination levels*. Available: https://ourlandandwater.nz/wp-content/uploads/2022/11/Output-ID-1808-OLW-report-MfE-2021-FINAL-ACookson-280621.pdf [Accessed 11/08/2025].

Morel-Journel, T., Lehtinen, S., Cotto, O., Amia, R., Dion, S., Figueroa, C., Martinson, J.N.V., Ralaimazava, P., Clermont, O., Duval, X., Nowrouzian, F.L., Walk, S.T., Denamur, E. and Blanquart, F. (2025) Residence-colonization trade-off and niche differentiation enable coexistence of *Escherichia coli* phylogroups in healthy humans. *The ISME Journal,* 19. Available at: https://doi.org/10.1093/ismejo/wraf089

Møretrø, T., Moen, B., Heir, E., Hansen, A.Å. and Langsrud, S. (2016) Contamination of salmon fillets and processing plants with spoilage bacteria. *International Journal of Food Microbiology,* 237**,** 98-108. Available at: https://doi.org/10.1016/j.ijfoodmicro.2016.08.016

Morgan, G., Pinchbeck, G., Haldenby, S., Schmidt, V. and Williams, N. (2024) Raw meat diets are a major risk factor for carriage of third-generation cephalosporin-resistant and multidrug-resistant *E. coli* by dogs in the UK. *Frontiers in Microbiology,* 15**,** 1460143. Available at: https://doi.org/10.3389/fmicb.2024.1460143

Muloi, D.M., Wee, B.A., McClean, D.M.H., Ward, M.J., Pankhurst, L., Phan, H., Ivens, A.C., Kivali, V., Kiyong'a, A., Ndinda, C., Gitahi, N., Ouko, T., Hassell, J.M., Imboma, T., Akoko, J., Murungi, M.K., Njoroge, S.M., Muinde, P., Nakamura, Y., Alumasa, L., Furmaga, E., Kaitho, T., Öhgren, E.M., Amanya, F., Ogendo, A., Wilson, D.J., Bettridge, J.M., Kiiru, J., Kyobutungi, C., Tacoli, C., Kang'ethe, E.K., Davila, J.D., Kariuki, S., Robinson, T.P., Rushton, J., Woolhouse, M.E.J. and Fèvre, E.M. (2022) Population genomics of *Escherichia coli* in livestock-keeping households across a rapidly developing urban landscape. *Nature Microbiology,* 7**,** 581-589. Available at: https://doi.org/10.1038/s41564-022-01079-y

Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Aguilar, G.R., Gray, A., Han, C., Bisignano, C., Rao, P. and Wool, E. (2022) Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet,* 399**,** 629-655. Available at: https://doi.org/10.1016/S0140-6736(21)02724-0

Nakamura, A., Takahashi, H., Arai, M., Tsuchiya, T., Wada, S., Fujimoto, Y., Shimabara, Y., Kuda, T. and Kimura, B. (2021) Molecular subtyping for source tracking of *Escherichia coli* using core genome multilocus sequence typing at a food manufacturing plant. *PLOS ONE,* 16**,** e0261352. Available at: https://doi.org/10.1371/journal.pone.0261352

Nascimento, F.F., dos Reis, M. and Yang, Z.H. (2017) A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution,* 1**,** 1446-1454. Available at: https://doi.org/10.1038/s41559-017-0280-x

National Center for Biotechnology Information (NCBI). (1988) *National Center for Biotechnology Information (NCBI)*. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. Available: https://www.ncbi.nlm.nih.gov/ [Accessed 26/03/2025].

Naziri, Z., Derakhshandeh, A., Firouzi, R., Motamedifar, M. and Shojaee Tabrizi, A. (2016) DNA fingerprinting approaches to trace *Escherichia coli* sharing between dogs and owners. *Journal of Applied Microbiology,* 120**,** 460-468. Available at: https://doi.org/10.1111/jam.13003

Naziri, Z., Poormaleknia, M. and Ghaedi Oliyaei, A. (2022) Risk of sharing resistant bacteria and/or resistance elements between dogs and their owners. *BMC Veterinary Research,* 18**,** 203. Available at: https://doi.org/10.1186/s12917-022-03298-1

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution,* 32**,** 268-274. Available at: https://doi.org/10.1093/molbev/msu300

Nicholls, H., Kainth, B., Cook, P., Mumford, R., Keevil, C., La Ragione, R., Threlfall, E., McDowell, D. and Callaghan, K. (2023) Foodborne antimicrobial resistance (AMR) research and surveillance in the UK: priorities identified through the Food Standards Agency's AMR programme review (2023). *Journal of medical microbiology,* 72. Available at: https://doi.org/10.1099/jmm.0.001753

Nielsen, K.L., Dynesen, P., Larsen, P. and Frimodt-Møller, N. (2014) Faecal *Escherichia coli* from patients with *E. coli* urinary tract infection and healthy controls who have never had a urinary tract infection. *Journal of Medical Microbiology,* 63**,** 582-589. Available at: https://doi.org/10.1099/jmm.0.068783-0

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B. and Ako-Adjei, D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research,* 44**,** D733-D745. Available at: https://doi.org/10.1093/nar/gkv1189

Ochman, H. and Selander, R.K. (1984) Standard reference strains of *Escherichia coli* from natural populations. *Journal of Bacteriology,* 157**,** 690-693. Available at: https://doi.org/10.1128/jb.157.2.690-693.1984

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S. and Phillippy, A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology,* 17. Available at: https://doi.org/10.1186/s13059-016-0997-x

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A. and Parkhill, J. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3693. Available at: https://doi.org/10.1093/bioinformatics/btv421

Palau, R., Bloomfield, S.J., Jenkins, C., Greig, D.R., Jorgensen, F. and Mather, A.E. (2024) *Yersinia enterocolitica* biovar 1A: An underappreciated potential pathogen in the food chain. *International Journal of Food Microbiology,* 412**,** 110554. Available at: https://doi.org/10.1016/j.ijfoodmicro.2023.110554

Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics,* 35**,** 526-528. Available at: https://doi.org/10.1093/bioinformatics/bty633

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research,* 25**,** 1043-55. Available at: https://doi.org/10.1101/gr.186072.114

Patkowski, J.B., Dahlberg, T., Amin, H., Gahlot, D.K., Vijayrajratnam, S., Vogel, J.P., Francis, M.S., Baker, J.L., Andersson, M. and Costa, T.R.D. (2023) The F-pilus biomechanical adaptability accelerates conjugative dissemination of antimicrobial resistance and biofilm formation. *Nature Communications,* 14**,** 1879. Available at: https://doi.org/10.1038/s41467-023-37600-y

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J. and Denamur, E. (1999) The Link between Phylogeny and Virulence in *Escherichia coli* Extraintestinal Infection. *Infection and Immunity,* 67**,** 546-553. Available at: https://doi.org/10.1128/IAI.67.2.546-553.1999

Pitout, J.D.D. and Chen, L. (2023) The Significance of Epidemic Plasmids in the Success of Multidrug-Resistant Drug Pandemic Extraintestinal Pathogenic *Escherichia coli*. *Infectious Disease and Therapy,* 12**,** 1029-1041. Available at: https://doi.org/10.1007/s40121-023-00791-4

Pitout, J.D.D., Peirano, G., Chen, L., DeVinney, R. and Matsumura, Y. (2022) *Escherichia coli* ST1193: following in the footsteps of *E. coli* ST131. *Antimicrobial Agents and Chemotherapy,* 66**,** e00511-22. Available at: https://doi.org/10.1128/aac.00511-22

Pokharel, P., Dhakal, S. and Dozois, C.M. (2023) The Diversity of *Escherichia coli* Pathotypes and Vaccination Strategies against This Versatile Bacterial Pathogen. *Microorganisms,* 11**,** 344. Available at: https://doi.org/10.3390/microorganisms11020344

Pomba, C., da Fonseca Joana, D., Baptista Bruno, C., Correia José, D. and Martínez-Martínez, L. (2009) Detection of the Pandemic O25-ST131 Human Virulent *Escherichia coli* CTX-M-15-Producing Clone Harboring the qnrB2 and aac(6')-Ib-cr Genes in a Dog. *Antimicrobial Agents and Chemotherapy,* 53**,** 327-328. Available at: https://doi.org/10.1128/aac.00896-08

Pomba, C., López-Cerero, L., Bellido, M., Serrano, L., Belas, A., Couto, N., Cavaco-Silva, P., Rodríguez-Baño, J. and Pascual, A. (2014) Within-lineage variability of ST131 *Escherichia coli* isolates from humans and companion animals in the south of Europe. *Journal of Antimicrobial Chemotherapy,* 69**,** 271-3. Available at: https://doi.org/10.1093/jac/dkt343

Pongrácz, P. and Dobos, P. (2023) What is a companion animal? An ethological approach based on Tinbergen's four questions. Critical review. *Applied Animal Behaviour Science,* 267**,** 106055. Available at: https://doi.org/10.1016/j.applanim.2023.106055

Price, M.N., Dehal, P.S. and Arkin, A.P. (2008) Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biology,* 9**,** R4. Available at: https://doi.org/10.1186/gb-2008-9-1-r4

Public Health (Control of Disease) Act 1984 c.22 (1984) [Accessed: 11/06/2025]

Public Health England. (2007) *Urinary tract infection: diagnostic tools for primary care*. GOV.UK: Public Health England,. Available: https://www.gov.uk/government/publications/urinary-tract-infection-diagnosis [Accessed 25/03/2025].

Public Health England. (2016) *Guidelines for the Investigation of Zoonotic Disease in England and Wales*. Available: https://www.gov.uk/government/publications/zoonotic-diseases-investigation-guidelines [Accessed 28/03/2025].

R Core Team. (2022) *R: a language and environment for statistical computing.* Foundation for Statistical Computing, Vienna, Austria. Available: https://www.R-project.org/ [Accessed 15 Apr 2025].

Randall, L.P., Horton, R.H., Chanter, J.I., Lemma, F. and Evans, S.J. (2021) A decline in the occurrence of extended-spectrum β-lactamase-producing *Escherichia coli* in retail chicken meat in the UK between 2013 and 2018. *Journal of Applied Microbiology,* 130**,** 247-257. Available at: https://doi.org/10.1111/jam.14687

Rasko, D.A., Rosovitz, M.J., Myers, G.S.A., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., Henderson, I.R., Sperandio, V. and Ravel, J. (2008) The pangenome structure of *Escherichia coli:* Comparative genomic analysis of *E-coli* commensal and pathogenic isolates. *Journal of Bacteriology,* 190**,** 6881-6893. Available at: https://doi.org/10.1128/JB.00619-08

Ratajczak, M., Laroche, E., Berthe, T., Clermont, O., Pawlak, B., Denamur, E. and Petit, F. (2010) Influence of hydrological conditions on the *Escherichia coli* population structure in the water of a creek on a rural watershed. *BMC Microbiology,* 10**,** 222. Available at: https://doi.org/10.1186/1471-2180-10-222

Ratnam, S., March, S.B., Ahmed, R., Bezanson, G.S. and Kasatiya, S. (1988) Characterization of *Escherichia coli* serotype O157:H7. *Journal of Clinical Microbiology,* 26**,** 2006-12. Available at: https://doi.org/10.1128/jcm.26.10.2006-2012.1988

Reid, C.J., Blau, K., Jechalke, S., Smalla, K. and Djordjevic, S.P. (2020) Whole genome sequencing of *Escherichia coli* from store-bought produce. *Frontiers in Microbiology,* 10. Available at: https://doi.org/10.3389/fmicb.2019.03050

Reisner, A., Höller, B.M., Molin, S. and Zechner, E.L. (2006) Synergistic effects in mixed *Escherichia coli* biofilms: conjugative plasmid transfer drives biofilm expansion. *Journal of bacteriology,* 188**,** 3582-8. Available at: https://doi.org/10.1128/JB.188.10.3582-3588.2006

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution,* 3**,** 217-223. Available at: https://doi.org/10.1111/j.2041-210X.2011.00169.x

Robins-Browne, R.M. (1987) Traditional Enteropathogenic *Escherichia coli* of Infantile Diarrhea. *Reviews of Infectious Diseases,* 9**,** 28-53. Available at: http://www.jstor.org/stable/4454034.

Robinson, D.F. and Foulds, L.R. (1979) Comparison of weighted labelled trees. *In:* Horadam, A.F. and Wallis, W.D. (ed. *Combinatorial Mathematics VI.* Springer, Berlin, Heidelberg.Available at: https://doi.org/10.1007/BFb0102690

Robinson, D.F. and Foulds, L.R. (1981) Comparison of phylogenetic trees. *Mathematical Biosciences,* 53**,** 131-147. Available at: https://doi.org/10.1016/0025-5564(81)90043-2

Rode, C.K., Melkerson-Watson, L.J., Johnson, A.T. and Bloch, C.A. (1999) Type-specific contributions to chromosome size differences in *Escherichia coli*. *Infect Immun,* 67**,** 230-6. Available at: https://doi.org/10.1128/IAI.67.1.230-236.1999

Rodrigues, C., Machado, E., Fernandes, S., Peixe, L. and Novais, Â. (2016) An update on faecal carriage of ESBL-producing Enterobacteriaceae by Portuguese healthy humans: detection of the H30 subclone of B2-ST131 *Escherichia coli* producing CTX-

M-27. *Journal of Antimicrobial Chemotherapy,* 71**,** 1120-1122. Available at: https://doi.org/10.1093/jac/dkv443

Rodríguez-Ezpeleta, N. and Philippe, H. (2009) Phylogenomics. *In:* Schaechter, Moselio (ed. *Encyclopedia of Microbiology (Third Edition).* Oxford: Academic Press.633-643. Available at: https://doi.org/10.1016/B978-012373944-5.00013-4

Roer, L., Tchesnokova, V., Allesøe, R., Muradova, M., Chattopadhyay, S., Ahrenfeldt, J., Thomsen, M.C.F., Lund, O., Hansen, F., Hammerum, A.M., Sokurenko, E. and Hasman, H. (2017) Development of a web tool for *Escherichia coli* subtyping based on *fimH* alleles. *Journal of Clinical Microbiology,* 55**,** 2538-2543. Available at: https://doi.org/10.1128/JCM.00737-17

Röhling, S., Linne, A., Schellhorn, J., Hosseini, M., Dencker, T. and Morgenstern, B. (2020) The number of k-mer matches between two DNA sequences as a function of k and applications to estimate phylogenetic distances. *PLOS ONE,* 15**,** e0228070. Available at: https://doi.org/10.1371/journal.pone.0228070

Ronco, T., Stegger, M., Olsen, R.H., Sekse, C., Nordstoga, A.B., Pohjanvirta, T., Lilje, B., Lyhs, U., Andersen, P.S. and Pedersen, K. (2017) Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. *BMC genomics,* 18. Available at: https://doi.org/10.1186/s12864-016-3415-6

Roschanski, N., Guenther, S., Vu, T.T.T., Fischer, J., Semmler, T., Huehn, S., Alter, T. and Roesler, U. (2017) VIM-1 carbapenemase-producing *Escherichia coli* isolated from retail seafood, Germany 2016. *Eurosurveillance,* 22**,** 17-00032. Available at: https://doi.org/10.2807/1560-7917.ES.2017.22.43.17-00032

Rosen, A. and Levin, R.E. (1970) Detection and Incidence of *Escherichia coli* on Storage Pen Surfaces of Fishing Trawlers. *Applied Microbiology,* 20**,** 103-106. Available at: https://doi.org/10.1128/am.20.1.103-106.1970

Rouger, A., Tresse, O. and Zagorec, M. (2017) Bacterial contaminants of poultry meat: sources, species, and dynamics. *Microorganisms,* 5**,** 50. Available at: https://doi.org/10.3390/microorganisms5030050

Royal Society for the Prevention of Cruelty to Animals. (2022) *Welfare of Meat Chickens*. Available: https://www.rspca.org.uk/documents/1494939/7712578/FAD-Meat-Chickens-Information-Sheet-2022.pdf/78c5de7e-a4e0-90cf-f75a-22e27baa81af?t=1673621417579 [Accessed].

RStudio Team. (2020) *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA. Available: http://www.rstudio.com/ [Accessed 15/04/2025].

Russo, T.A. and Johnson, J.R. (2000) Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of *Escherichia coli*: ExPEC. *The Journal of Infectious Diseases,* 181**,** 1753-1754. Available at: https://doi.org/10.1086/315418

Sabate, M. and Prats, G. (2002) Structure and function of integrons. *Enfermedades Infecciosas Y Microbiologia Clinica,* 20**,** 341-345. Available at: https://doi.org/10.1016/s0213-005x(02)72813-9

Saida, H., Ytow, N. and Seki, H. (1998) Photometric application of the Gram stain method to characterize natural bacterial populations in aquatic environments. *Applied and Environmental Microbiology,* 64**,** 742-747. Available at: https://doi.org/10.1128/AEM.64.2.742-747.1998

Sakoparnig, T., Field, C. and van Nimwegen, E. (2021) Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. *Elife,* 10. Available at: https://doi.org/10.7554/eLife.65366

Santos MI, Cavaco A., Gouveia J., Novais M.R., Nogueira P.J., Pedroso L. and M.A.S.S., F. (2012) Evaluation of minimally processed salads commercialized in Portugal. *Food Control,* 23**,** 275-281. Available at: https://doi.org/10.1016/j.foodcont.2011.06.022

Saria, P., Doulidis, P.G., Desvars-Larrive, A., Tóth, A.G., Burgener, I.A., Rodríguez-Rojas, A. and Makarova, O. (2025) Pandemic human-associated extended-spectrum β-lactamase-producing *Escherichia coli* lineages of ST38, ST131 and ST141 identified in Viennese dogs. *Journal of Antimicrobial Chemotherapy***,** dkaf103. Available at: https://doi.org/10.1093/jac/dkaf103

Schmidt, V.M., Pinchbeck, G.L., Nuttall, T., McEwan, N., Dawson, S. and Williams, N.J. (2015) Antimicrobial resistance risk factors and characterisation of faecal *E. coli* isolated from healthy Labrador retrievers in the United Kingdom. *Preventive Veterinary Medicine,* 119**,** 31-40. Available at: https://doi.org/10.1016/j.prevetmed.2015.01.013

Schmidt, W. (1995) Phylogeny reconstruction for protein sequences based on amino acid properties. *Journal of Molecular Evolution,* 41**,** 522-530. Available at: https://doi.org/10.1007/BF00160324

Schwengers, O., Hain, T., Chakraborty, T. and Goesmann, A. (2020) ReferenceSeeker: rapid determination of appropriate reference genomes. *Journal of Open Source Software,* 5**,** 1994. Available at: https://doi.org/10.21105/joss.01994

Sealey, J.E., Hammond, A., Reyher, K.K. and Avison, M.B. (2023) One health transmission of fluoroquinolone-resistant *Escherichia coli* and risk factors for their excretion by dogs living in urban and nearby rural settings. *One Health,* 17**,** 100640. Available at: https://doi.org/10.1016/j.onehlt.2023.100640

Sears, H.J. and Brownlee, I. (1952) Further observations on the persistence of individual strains of *Escherichia coli* in the intestinal tract of man. *Journal of bacteriology,* 63**,** 47-57. Available at: https://doi.org/10.1128/jb.63.1.47-57.1952

Sears, H.J., Brownlee, I. and Uchiyama, J.K. (1950) Persistence of individual strains of *Escherichia coli* in the intestinal tract of man. *Journal of Bacteriology,* 59**,** 293-301. Available at: https://doi.org/10.1128/jb.59.2.293-301.1950

Seemann, T. (2015) *Snippy: fast bacterial variant calling from NGS reads*. Available: https://github.com/tseemann/snippy [Accessed 26 Aug 2025].

Seemann, T. (2022) *MLST*. Available: https://github.com/tseemann/mlst [Accessed 13/04/2022].

Seemann, T., Klötzl, F. and Terhorst, J. (2021) *snp-dists*. Available: https://github.com/tseemann/snp-dists [Accessed 26 Aug 2025].

Segerman, B. (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in Cellular and Infection Microbiology,* 2. Available at: https://doi.org/10.3389/fcimb.2012.00116

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N. and Whittam, T.S. (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and environmental microbiology,* 51**,** 873-84. Available at: https://doi.org/10.1128/aem.51.5.873-884.1986

Selander, R.K. and Levin, B.R. (1980) Genetic diversity and structure in *Escherichia coli* populations. *Science,* 210**,** 545-547. Available at: https://doi.org/10.1126/science.6999623

Semple, C. and Steel, M.A. (2003) Phylogenetics. *Oxford lecture series in mathematics and its applications ; 24.* Oxford: Oxford University Press

Shakya, M., Ahmed, S.A., Davenport, K.W., Flynn, M.C., Lo, C.C. and Chain, P.S.G. (2020) Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. *Scientific Reports,* 10. Available at: https://doi.org/10.1038/s41598-020-58356-1

Shiomi, D., Mori, H. and Niki, H. (2009) Genetic mechanism regulating bacterial cell shape and metabolism. *Communicative & integrative biology,* 2**,** 219-20. Available at: https://doi.org/10.4161/cib.2.3.7930

Simonsen, M., Mailund, T. and Pedersen, C.N.S. (2008) Rapid neighbour-joining. *Algorithms in Bioinformatics,* 5251**,** 113-122. Available at: https://doi.org/10.1007/978-3-540-87361-7_10

Sims, G.E. and Kim, S.-H. (2011) Whole-genome phylogeny of *Escherichia coli/Shigella* group by feature frequency profiles (FFPs). *Proceedings of the National Academy of Sciences,* 108**,** 8329-8334. Available at: https://doi.org/10.1073/pnas.1105168108

Skippington, E. and Ragan, M.A. (2012) Phylogeny rather than ecology or lifestyle biases the construction of Escherichia coli-Shigella genetic exchange communities. *Open Biology,* 2**,** 120112. Available at: https://doi.org/10.1098/rsob.120112

Slatkin, M. (2008) Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics,* 9**,** 477-485. Available at: https://doi.org/10.1038/nrg2361

Sola, M., Mani, Y., Saras, E., Drapeau, A., Grami, R., Aouni, M., Madec, J.-Y., Haenni, M. and Mansour, W. (2022) Prevalence and Characterization of Extended-Spectrum β-Lactamase- and Carbapenemase-Producing Enterobacterales from Tunisian Seafood. *Microorganisms,* 10**,** 1364. Available at: https://doi.org/10.3390/microorganisms10071364

Sonda, T., Kumburu, H., van Zwetselaar, M., Alifrangis, M., Mmbaga, B.T., Aarestrup, F.M., Kibiki, G. and Lund, O. (2018) Whole genome sequencing reveals high clonal diversity of *Escherichia coli* isolated from patients in a tertiary care hospital in Moshi, Tanzania. *Antimicrobial Resistance & Infection Control,* 7**,** 1-12. Available at: https://doi.org/10.1186/s13756-018-0361-x

Steel, M.A. and Penny, D. (1993) Distributions of Tree Comparison Metrics—Some New Results. *Systematic Biology,* 42**,** 126-141. Available at: https://doi.org/10.1093/sysbio/42.2.126

Steenwyk, J.L., Li, Y., Zhou, X., Shen, X.-X. and Rokas, A. (2023) Incongruence in the phylogenomics era. *Nature Reviews Genetics*. Available at: https://doi.org/10.1038/s41576-023-00620-x

Stege Paul, B., Hordijk, J., Sandholt Arnar, K.S., Zomer Aldert, L., Viveen Marco, C., Rogers Malbert, R.C., Salomons, M., Wagenaar Jaap, A., Mughini-Gras, L., Willems Rob, J.L. and Paganelli Fernanda, L. (2023) Gut Colonization by ESBL-Producing *Escherichia coli* in Dogs Is Associated with a Distinct Microbiome and Resistome Composition. *Microbiology Spectrum,* 11**,** e00063-23. Available at: https://doi.org/10.1128/spectrum.00063-23

Stenske, K.A., Bemis, D.A., Gillespie, B.E., D'Souza, D.H., Oliver, S.P., Draughon, F.A., Matteson, K.J. and Bartges, J.W. (2009) Comparison of clonal relatedness and antimicrobial susceptibility of fecal *Escherichia coli* from healthy dogs and their owners. *American Journal of Veterinary Research,* 70**,** 1108-1116. Available at: https://doi.org/10.2460/ajvr.70.9.1108

Stoesser, N., Sheppard, A.E., Moore, C., Golubchik, T., Parry, C.M., Nget, P., Saroeun, M., Day, N.P., Giess, A. and Johnson, J. (2015) Extensive within-host diversity in fecally carried extended-spectrum-beta-lactamase-producing *Escherichia coli* isolates: implications for transmission analyses. *Journal of clinical microbiology,* 53**,** 2122-2131. Available at: https://doi.org/10.1128/JCM.00378-15

Sung, K., Park, M., Kweon, O., Paredes, A., Savenka, A. and Khan, S.A. (2025) Proteomic insights into dual-species biofilm formation of *E. coli* and *E. faecalis* on urinary catheters. *Scientific Reports,* 15**,** 3739. Available at: https://doi.org/10.1038/s41598-024-81953-3

Tam, C.C., Rodrigues, L.C., Viviani, L., Dodds, J.P., Evans, M.R., Hunter, P.R., Gray, J.J., Letley, L.H., Rait, G., Tompkins, D.S. and Brien, S.J. (2012) Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut,* 61**,** 69. Available at: https://doi.org/10.1136/gut.2011.238386

Tartof, S.Y., Solberg, O.D., Manges, A.R. and Riley, L.W. (2005) Analysis of a Uropathogenic *Escherichia coli* Clonal Group by Multilocus Sequence Typing. *Journal of Clinical Microbiology,* 43**,** 5860-5864. Available at: https://doi.org/10.1128/JCM.43.12.5860-5864.2005

Tawfick, M.M., Elshamy, A.A., Mohamed, K.T. and El Menofy, N.G. (2022) Gut Commensal Escherichia coli, a High-Risk Reservoir of Transferable Plasmid-Mediated Antimicrobial Resistance Traits. *Infection and drug resistance,* 15**,** 1077-1091. Available at: https://doi.org/10.2147/IDR.S354884

Tenaillon, O., Skurnik, D., Picard, B. and Denamur, E. (2010) The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology,* 8**,** 207-217. Available at: https://doi.org/10.1038/nrmicro2298

Ternent, H.E., Innocent, G.T., Filshie, L.M., Taylor, D.J., Steele, W.B., McEwen, S.A., Reilly, W.J., Gunn, G.J., Reid, S.W. and Mellor, D.J. (2004) Frozen storage of *Escherichia coli* O157 in buffered peptone water and its detection on bovine carcasses. *Journal of Food Protection,* 67**,** 40-5. Available at: https://doi.org/10.4315/0362-028X-67.1.40

Timofte, D., Maciuca, I.E., Kemmett, K., Wattret, A. and Williams, N.J. (2014) Detection of the human-pandemic *Escherichia coli* B2-O25b-ST131 in UK dogs. *Veterinary Record,* 174**,** 352-352. Available at: https://doi.org/10.1136/vr.101893

Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W. and Corander, J. (2018) RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome open research,* 3**,** 93-93. Available at: https://doi.org/10.12688/wellcomeopenres.14694.1

Toombs-Ruane, L.J., Benschop, J., French, N.P., Biggs, P.J., Midwinter, A.C., Marshall, J.C., Chan, M., Drinković, D., Fayaz, A., Baker, M.G., Douwes, J., Roberts, M.G. and Burgess, S.A. (2020) Carriage of extended-spectrum-beta-lactamase- and AmpC beta-lactamase-producing *Escherichia coli* strains from humans and pets in the same households. *Applied and Environmental Microbiology,* 86**,** e01613-20. Available at: https://doi.org/10.1128/AEM.01613-20

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., El Karoui, M., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Le Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Saint Ruf, C., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Medigue, C., Rocha, E.P.C. and Denamur, E. (2009) Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. *Plos Genetics,* 5. Available at: https://doi.org/10.1371/journal.pgen.1000344

Touchon, M., Perrin, A., de Sousa, J.A.M., Vangchhia, B., Burn, S., O'Brien, C.L., Denamur, E., Gordon, D. and Rocha, E.P.C. (2020) Phylogenetic background and habitat drive the genetic diversification of Escherichia coli. *PLOS Genetics,* 16**,** e1008866. Available at: https://doi.org/10.1371/journal.pgen.1008866

UK Health Security Agency. (2023a) *2023 to 2024 report of UK national reference laboratory for food microbiology*. UK Health Security Agency. Available: https://www.gov.uk/government/publications/national-reference-laboratory-for-food-microbiology-annual-report [Accessed 11/08/2025].

UK Health Security Agency. (2023b) *Shiga toxin-producing Escherichia coli: public health management*. Available: https://www.gov.uk/government/publications/shiga-toxin-producing-escherichia-coli-public-health-management [Accessed 11/08/2025].

UK Health Security Agency. (2024) *English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) Report 2023 to 2024*. Available: https://assets.publishing.service.gov.uk/media/6734e208b613efc3f1823095/ESPAUR-report-2023-2024.pdf [Accessed 11/06/2025].

UK Health Security Agency. (2025a) *Quarterly epidemiological commentary: mandatory Gram-negative bacteraemia, MRSA, MSSA and C. difficile infections (data up to July to September 2024)*. GOV.UK: UK Health Security Agency,. Available: https://www.gov.uk/government/statistics/mrsa-mssa-gram-negative-bacteraemia-and-cdi-quarterly-report/quarterly-epidemiological-commentary-mandatory-gram-negative-bacteraemia-mrsa-mssa-and-c-difficile-infections-data-up-to-july-to-september-2024 [Accessed].

UK Health Security Agency. (2025b) *UK Access, Watch, Reserve, and Other classification for antibiotics (UK-AWaRe antibiotic classification)*. GOV.UK. Available: https://www.gov.uk/government/publications/uk-aware-antibiotic-classification/uk-access-watch-reserve-and-other-classification-for-antibiotics-uk-aware-antibiotic-classification [Accessed 03/06/2025].

UK Health Security Agency and Veterinary Medicines Directorate. (2023) *Executive summary Third UK One Health Report*. Available: https://www.gov.uk/government/publications/uk-one-health-report-joint-report-on-antibiotic-use-antibiotic-sales-and-antibiotic-resistance/executive-summary-third-uk-one-health-report#:~:text=Sales%20data%20on%20veterinary%20antibiotics,potential%20for%20transfer%20of%20AMR. [Accessed 04/04/2025].

UK Pet Food. (2024) *UK pet population*. Available: https://www.ukpetfood.org/information-centre/statistics/uk-pet-population.html [Accessed 24 Sep 2024].

Ulitsky, I., Burstein, D., Tuller, T. and Chor, B. (2006) The Average Common Substring Approach to Phylogenomic Reconstruction. *Journal of Computational Biology,* 13**,** 336-350. Available at: https://doi.org/10.1089/cmb.2006.13.336

Valenza, G., Werner, M., Eisenberger, D., Nickel, S., Lehner-Reindl, V., Höller, C. and Bogdan, C. (2019) First report of the new emerging global clone ST1193 among clinical isolates of extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* from Germany. *Journal of Global Antimicrobial Resistance,* 17**,** 305-308. Available at: https://doi.org/10.1016/j.jgar.2019.01.014

van Bree, F.P.J., Bokken, G., Mineur, R., Franssen, F., Opsteegh, M., van der Giessen, J.W.B., Lipman, L.J.A. and Overgaauw, P.A.M. (2018) Zoonotic bacteria and parasites found in raw meat-based diets for cats and dogs. *The Veterinary record,* 182**,** 50. Available at: https://doi.org/10.1136/vr.104535

van den Bunt, G., Fluit, A.C., Spaninks, M.P., Timmerman, A.J., Geurts, Y., Kant, A., Scharringa, J., Mevius, D., Wagenaar, J.A., Bonten, M.J.M., van Pelt, W. and Hordijk, J. (2020) Faecal carriage, risk factors, acquisition and persistence of ESBL-producing Enterobacteriaceae in dogs and cats and co-carriage with humans belonging to the same household. *Journal of Antimicrobial Chemotherapy,* 75**,** 342-350. Available at: https://doi.org/10.1093/jac/dkz462

van der Putten, B.C.L., Matamoros, S., Mende, D.R., Scholl, E.R., consortium†, C. and Schultsz, C. (2021) Escherichia ruysiae sp. nov., a novel Gram-stain-negative bacterium, isolated from a faecal sample of an international traveller. *International Journal of Systematic and Evolutionary Microbiology,* 71. Available at: https://doi.org/10.1099/ijsem.0.004609

Vangchhia, B., Blyton, M.D.J., Collignon, P., Kennedy, K. and Gordon, D.M. (2018) Factors affecting the presence, genetic diversity and antimicrobial sensitivity of *Escherichia coli* in poultry meat samples collected from Canberra, Australia. *Environmental microbiology,* 20**,** 1350-1361. Available at: https://doi.org/10.1111/1462-2920.14030

Veterinary Medicines Directorate. (2024) *UK veterinary antibiotic resistance and sales surveillance report UK-VARSS 2023*. Available: https://www.gov.uk/government/publications/veterinary-antimicrobial-resistance-and-sales-surveillance-2023 [Accessed 25/09/2025].

Veterinary Medicines Directorate. (2025) *UK launches world-first study to assess antibiotic resistance levels in healthy dogs and cats*. Available: https://www.gov.uk/government/news/uk-launches-world-first-study-to-assess-antibiotic-resistance-levels-in-healthy-dogs-and-cats [Accessed 22/09/2025].

Vingino, A., Roberts, M.C., Wainstein, M., West, J., Norman, S.A., Lambourn, D., Lahti, J., Ruiz, R., D'Angeli, M. and Weissman, S.J. (2021) Surveillance for antibiotic-resistant *E. coli* in the Salish Sea Ecosystem. *Antibiotics,* 10**,** 1201. Available at: https://doi.org/10.3390/antibiotics10101201

Vogan, A.A. and Higgs, P.G. (2011) The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biology Direct,* 6. Available at: https://doi.org/10.1186/1745-6150-6-1

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M. and Whittam, T.S. (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environmental Microbiology,* 9**,** 2274-2288. Available at: https://doi.org/10.1111/j.1462-2920.2007.01341.x

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M. and Whittam, T.S. (2009) Cryptic Lineages of the Genus *Escherichia*. *Applied and Environmental Microbiology,* 75**,** 6534-6544. Available at: https://doi.org/10.1128/AEM.01262-09

Wang, Q., Han, Y.Y., Zhang, T.J., Chen, X., Lin, H., Wang, H.N. and Lei, C.W. (2024) Whole-genome sequencing of *Escherichia coli* from retail meat in China reveals the dissemination of clinically important antimicrobial resistance genes. *International Journal of Food Microbiology,* 415**,** 110634. Available at: https://doi.org/10.1016/j.ijfoodmicro.2024.110634

Wang, X., Jordan, I.K. and Mayer, L.W. (2015) Chapter 29 - A Phylogenetic Perspective on Molecular Epidemiology. *In:* Tang, Yi-Wei, Sussman, Max, Liu, Dongyou, Poxton, Ian and Schwartzman, Joseph (ed. 2) *Molecular Medical Microbiology (Second*

Edition). Boston: Academic Press.517-536. Available at: https://doi.org/10.1016/B978-0-12-397169-2.00029-9

Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature,* 171**,** 737-738. Available at: https://doi.org/10.1038/171737a0

Wensley, S., Betton, V., Gosschalk, K., Hooker, R., Main, D.C.J., Martin, N. and Tipton, E. (2021) Driving evidence-based improvements for the UK's ‘Stressed. Lonely. Overweight. Bored. Aggressive. Misunderstood...but loved’ companion animals. *Veterinary Record,* 189**,** e7. Available at: https://doi.org/10.1002/vetr.7

White, R.T., Bull, M.J., Barker, C.R., Arnott, J.M., Wootton, M., Jones, L.S., Howe, R.A., Morgan, M., Ashcroft, M.M., Forde, B.M., Connor, T.R. and Beatson, S.A. (2024) Genomic epidemiology reveals geographical clustering of multidrug-resistant *Escherichia coli* ST131 associated with bacteraemia in Wales. *Nature Communications,* 15**,** 1371. Available at: https://doi.org/10.1038/s41467-024-45608-1

Whittam, T.S., Ochman, H. and Selander, R.K. (1983) Multilocus genetic structure in natural populations of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America,* 80**,** 1751-5. Available at: https://doi.org/10.1073/pnas.80.6.1751

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Available: https://ggplot2.tidyverse.org [Accessed 11/08/2025].

Williams, D., Fournier, G.P., Lapierre, P., Swithers, K.S., Green, A.G., Andam, C.P. and Gogarten, J.P. (2011) A Rooted Net of Life. *Biology Direct,* 6**,** 45. Available at: https://doi.org/10.1186/1745-6150-6-45

Wilson, D., Dolan, G., Aird, H., Sorrell, S., Dallman, T.J., Jenkins, C., Robertson, L. and Gorton, R. (2018) Farm-to-fork investigation of an outbreak of Shiga toxin-producing *Escherichia coli* O157. *Microbial Genomics,* 4. Available at: https://doi.org/10.1099/mgen.0.000160

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C., Ochman, H. and Achtman, M. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular microbiology,* 60**,** 1136-51. Available at: https://doi.org/10.1111/j.1365-2958.2006.05172.x

World Health Organization. (2015) *WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015*. Geneva: World Health Organization. Available: https://iris.who.int/handle/10665/199350 [Accessed 11/06/2025].

World Health Organization. (2019) *Critically important antimicrobials for human medicine: 6th revision*. World Health Organization. Available: https://www.who.int/groups/advisory-group-on-the-who-list-of-critically-important-antimicrobials [Accessed 11 Jun 2025].

World Health Organization. (2022) *WHO global strategy for food safety 2022-2030: towards stronger food safety systems and global cooperation*. Available: https://www.who.int/publications/i/item/9789240057685 [Accessed 11/06/2025].

World Health Organization. (2023) *AWaRe classification of antibiotics for evaluation and monitoring of use, 2023*. Available: https://www.who.int/publications/i/item/WHO-MHP-HPS-EML-2023.04 [Accessed 03/06/2025].

World Health Organization. (2024) *WHO medically important antimicrobial list*. Available: https://www.who.int/news/item/08-02-2024-who-medically-important-antimicrobial-list-2024 [Accessed 11 Jun 2025].

World Health Organization. (2025) *One Health*. Available: https://www.who.int/health-topics/one-health#tab=tab_1 [Accessed 20 March 2025].

Xia, F., Jiang, M., Wen, Z., Wang, Z., Wang, M., Xu, Y., Zhuge, X. and Dai, J. (2022) Complete genomic analysis of ST117 lineage extraintestinal pathogenic *Escherichia coli* (ExPEC) to reveal multiple genetic determinants to drive its global transmission: ST117 *E. coli* as an emerging multidrug-resistant foodborne ExPEC with zoonotic potential. *Transboundary and Emerging Diseases,* 69**,** 3256-3273. Available at: https://doi.org/10.1111/tbed.14678

Xu, J. (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular Ecology,* 15**,** 1713-31. Available at: https://doi.org/10.1111/j.1365-294X.2006.02882.x

Yamaji, R., Friedman, C.R., Rubin, J., Suh, J., Thys, E., McDermott, P., Hung-Fan, M. and Riley, L.W. (2018) A Population-Based Surveillance Study of Shared Genotypes of *Escherichia coli* Isolates from Retail Meat and Suspected Cases of Urinary Tract Infections. *Msphere,* 3. Available at: https://doi.org/10.1128/mSphere.00179-18

Yang, Z. (2014) Phylogeny reconstruction: overview. *In:* Yang, Ziheng (ed. *Molecular Evolution: A Statistical Approach.* Oxford University Press.0. Available at: https://doi.org/10.1093/acprof:oso/9780198567028.003.0003

Young, A.D. and Gillung, J.P. (2020) Phylogenomics - principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology,* 45**,** 225-247. Available at: https://doi.org/10.1111/syen.12406

Yu, D., Banting, G. and Neumann, N.F. (2021) A review of the taxonomy, genetics, and biology of the genus Escherichia and the type species *Escherichia coli*. *Canadian Journal of Microbiology,* 67**,** 553-571. Available at: https://doi.org/10.1139/cjm-2020-0508

Yu, G. (2020) Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics,* 69**,** e96. Available at: https://doi.org/10.1002/cpbi.96

Zankari, E., Allesøe, R., Joensen, K.G., Cavaco, L.M., Lund, O. and Aarestrup, F.M. (2017) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *Journal of Antimicrobial Chemotherapy,* 72**,** 2764-2768. Available at: https://doi.org/10.1093/jac/dkx217

Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics,* 19**,** 153. Available at: https://doi.org/10.1186/s12859-018-2129-y

Zhang, C., Scornavacca, C., Molloy, E.K. and Mirarab, S. (2020) ASTRAL-Pro: Quartet-Based Species-Tree Inference despite Paralogy. *Molecular Biology and Evolution,* 37**,** 3292-3307. Available at: https://doi.org/10.1093/molbev/msaa139

Zhang, W.-L., Bielaszewska, M., Bockemühl, J., Schmidt, H., Scheutz, F. and Karch, H. (2000) Molecular Analysis of H Antigens Reveals that Human Diarrheagenic *Escherichia coli* O26 Strains That Carry the *eae* Gene Belong to the H11 Clonal Complex. *Journal of Clinical Microbiology,* 38**,** 2989-2993. Available at: https://doi.org/10.1128/JCM.38.8.2989-2993.2000

Zhao, R., Hao, J., Yang, J., Tong, C., Xie, L., Xiao, D., Zeng, Z. and Xiong, W. (2022) The co-occurrence of antibiotic resistance genes between dogs and their owners in families. *Imeta,* 1**,** e21. Available at: https://doi.org/10.1002/imt2.21

Zhi, S., Banting, G., Stothard, P., Ashbolt, N.J., Checkley, S., Meyer, K., Otto, S. and Neumann, N.F. (2019) Evidence for the evolution, clonal expansion and global dissemination of water treatment-resistant naturalized strains of *Escherichia coli* in wastewater. *Water Research,* 156**,** 208-222. Available at: https://doi.org/10.1016/j.watres.2019.03.024

Zhou, Z., Alikhan, N.F., Mohamed, K., Fan, Y. and Achtman, M. (2020) The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome research,* 30**,** 138-152. Available at: https://doi.org/10.1101/gr.251678.119

Zhuge, X., Zhou, Z., Jiang, M., Wang, Z., Sun, Y., Tang, F., Xue, F., Ren, J. and Dai, J. (2021) Chicken-source *Escherichia coli* within phylogroup F shares virulence genotypes and is closely related to extraintestinal pathogenic *E. coli* causing human infections. *Transboundary and Emerging Diseases,* 68**,** 880-895. Available at: https://doi.org/10.1111/tbed.13755

Zielezinski, A., Vinga, S., Almeida, J. and Karlowski, W.M. (2017) Alignment-free sequence comparison: benefits, applications, and tools. *Genome Biology,* 18**,** 186. Available at: https://doi.org/10.1186/s13059-017-1319-7

Zwirzitz, B., Wetzels, S.U., Dixon, E.D., Stessl, B., Zaiser, A., Rabanser, I., Thalguter, S., Pinior, B., Roch, F.-F., Strachan, C., Zanghellini, J., Dzieciol, M., Wagner, M. and Selberherr, E. (2020) The sources and transmission routes of microbial populations throughout a meat processing facility. *npj Biofilms and Microbiomes,* 6**,** 26. Available at: https://doi.org/10.1038/s41522-020-0136-z

Appendices

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
36855171	Quinolone-resistant *Escherichia coli* at the interface between humans, poultry and their shared environment- a potential public health risk	2023	10.1186/s42522-023-00079-0	cgMLST
33671102	High Risk Clone: A Proposal of Criteria Adapted to the One Health Context with Application to Enterotoxigenic *Escherichia coli* in the Pig Population	2021	10.3390/antibiotics10030244	multiple alignment and Maximum likelihood tree
33757589	Genetic relatedness of multidrug resistant *Escherichia coli* isolated from humans, chickens and poultry environments	2021	10.1186/s13756-021-00930-x	Maximum likelihood SNP tree
34568479	Occurrence of Colibacillosis in Broilers and Its Relationship With Avian Pathogenic *Escherichia coli* (APEC) Population Structure and Molecular Characteristics	2021	10.3389/fvets.2021.737720	Maximum likelihood SNP tree
37104397	Description of Antimicrobial-Resistant *Escherichia coli* and Their Dissemination Mechanisms on Dairy Farms	2023	10.3390/vetsci10040242	cgMLST, SNP tree

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
37338386	Previous Antibiotic Exposure Reshapes the Population Structure of Infecting Uropathogenic *Escherichia coli* Strains by Selecting for Antibiotic Resistance over Urovirulence	2023	10.1128/spectrum.05242-22	Maximum likelihood SNP tree against reference
34777266	The Population Genetics, Virulence, and Public Health Concerns of *Escherichia coli* Collected From Rats Within an Urban Environment	2021	10.3389/fmicb.2021.631761	cgMLST alignment, SNPs extracted and substitution model, Maximum likelihood
37374960	*Escherichia coli* Strains Display Varying Susceptibility to Grazing by the Soil Amoeba *Dictyostelium discoideum*	2023	10.3390/microorganisms11061457	core gene alignment and RaxML
34878971	The global population structure and evolutionary history of the acquisition of major virulence factor-encoding genetic elements in Shiga toxin-producing *Escherichia coli* O121:H19	2021	10.1099/mgen.0.000716	Maximum likelihood SNP tree against reference
33627701	Population structure and genetic diversity of non-O157 Shiga toxin-producing *Escherichia coli* (STEC) clinical isolates from Michigan	2021	10.1038/s41598-021-83775-z	MLST sequence alignmnet and NJ

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
36480516	Effects of chronic exposure to arsenic on the fecal carriage of antibiotic-resistant *Escherichia coli* among people in rural Bangladesh	2022	10.1371/journal.ppat.1010952	rMLST and concatenated
35544167	Emergence and dissemination of antimicrobial resistance in *Escherichia coli* causing bloodstream infections in Norway in 2002-17: a nationwide, longitudinal, microbial population genomic study	2021	10.1016/S2666-5247(21)00031-8	Maximum likelihood SNP tree against reference
34047644	The Avian Pathogenic *Escherichia coli* (APEC) pathotype is comprised of multiple distinct, independent genotypes	2021	10.1080/03079457.2021.1915960	Core genome ParSNP
35392797	Rates of evolutionary change of resident *Escherichia coli* O157:H7 differ within the same ecological niche	2022	10.1186/s12864-022-08497-6	SNPs in a multiple sequence alignment file and NJ.
34935444	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	2021	10.1128/msphere.00868-21	core genome, maximum likelihood

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
33416498	Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species	2021	10.7554/eLife.65366	Core genome alignment and core alignment blocks
36406419	Non-lactose fermenting *Escherichia coli*: Following in the footsteps of lactose fermenting *E. coli* high-risk clones	2022	10.3389/fmicb.2022.1027494	alignment mapping to multiple reference sequences (REALPHY)
33408222	Genomic Epidemiology and Evolution of *Escherichia coli* in Wild Animals in Mexico	2021	10.1128/mSphere.00738-20	Roary, SNPs extracted and mapped against reference using an alignmnet free approach, similarity matrix and converted into distance matrix & NJ

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
33481909	The utility of *Escherichia coli* as a contamination indicator for rural drinking water: Evidence from whole genome sequencing	2021	10.1371/journal.pone.0245910	Roary and Fastree (approximate maximum-likelihood)
33952335	Phylogroup stability contrasts with high within sequence type complex dynamics of *Escherichia coli* bloodstream infection isolates over a 12-year period	2021	10.1186/s13073-021-00892-0	Roary, Iqtree
33961542	Insights into the acquisition of the pks island and production of colibactin in the *Escherichia coli* population	2021	10.1099/mgen.0.000579	Roary, snip-sites and RaxML
34431197	The E phylogroup of *Escherichia coli* is highly diverse and mimics the whole *E. coli* species population structure	2021	10.1111/1462-2920.15742	Roary Iqtree
34534282	Spread of ESC-, carbapenem- and colistin-resistant *Escherichia coli* clones and plasmids within and between food workers in Lebanon	2021	10.1093/jac/dkab327	Roary, gubbins, RaxML

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
34710106	Genomic analysis of shiga toxin-containing *Escherichia coli* O157:H7 isolated from Argentinean cattle	2021	10.1371/journal.pone.0258753	Roary, RAxML
35138158	Characterization of Genetic Elements Carrying mcr-1 Gene in *Escherichia coli* from the Community and Hospital Settings in Vietnam	2022	10.1128/spectrum.01356-21	Roary, Iqtree
35333870	Whole-genome sequencing and gene sharing network analysis powered by machine learning identifies antibiotic resistance sharing between animals, humans and environment in livestock farming	2022	10.1371/journal.pcbi.1010018	Roary, Iqtree
35945191	Complete genomic analysis of ST117 lineage extraintestinal pathogenic *Escherichia coli* (ExPEC) to reveal multiple genetic determinants to drive its global transmission: ST117 *E. coli* as an emerging multidrug-resistant foodborne ExPEC with zoonotic potential	2022	10.1111/tbed.14678	Roary, RAxML
36585742	A one-year genomic investigation of *Escherichia coli* epidemiology and nosocomial spread at a large US healthcare network	2022	10.1186/s13073-022-01150-7	Roary, SNP calling

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
36752777	Identification of genes influencing the evolution of *Escherichia coli* ST372 in dogs and humans	2023	10.1099/mgen.0.000930	Roary, Iqtree
36951916	Global population structure, genomic diversity and carbohydrate fermentation characteristics of clonal complex 119 (CC119), an understudied Shiga toxin-producing *E. coli* (STEC) lineage including O165:H25 and O172:H25	2023	10.1099/mgen.0.000959	Roary, SNP sites and RaxML
37355007	Stress Resistance and Virulence Gene Profiles Associated with Phylogeny and Phenotypes of *Escherichia coli* from Cattle	2023	10.1016/j.jfp.2023.100122	Roary, RAxML
37392822	Population genomics of diarrheagenic *Escherichia coli* uncovers high connectivity between urban and rural communities in Ecuador	2023	10.1016/j.meegid.2023.105476	Roary, Iqtree
32561945	Investigation of On-Farm Transmission Routes for Contamination of Dairy Cows with Top 7 *Escherichia coli* O-Serogroups	2021	10.1007/s00248-020-01542-5	compared to reference, SNP detection, presence, absence matrix

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
33232750	Extended-spectrum beta-lactamase-producing *Escherichia coli* and antimicrobial resistance in municipal and hospital wastewaters in Czech Republic: Culture-based and metagenomic approaches	2021	10.1016/j.envres.2020.110487	Maximum likelihood SNP tree
34431692	Arrangements of Mobile Genetic Elements among Virotype E Subpopulation of *Escherichia coli* Sequence Type 131 Strains with High Antimicrobial Resistance and Virulence Gene Content	2021	10.1128/mSphere.00550-21	Maximum likelihood SNP tree
34904942	Genomic insights into the circulation of pandemic fluoroquinolone-resistant extra-intestinal pathogenic *Escherichia coli* ST1193 in Vietnam	2021	10.1099/mgen.0.000733	Maximum likelihood SNP tree against reference
34671331	International High-Risk Clones Among Extended-Spectrum β-Lactamase-Producing *Escherichia coli* in Dhaka, Bangladesh	2021	10.3389/fmicb.2021.736464	Maximum likelihood SNP tree against reference
34910614	Genomic comparisons of *Escherichia coli* ST131 from Australia	2021	10.1099/mgen.0.000721	Maximum likelihood SNP tree against reference

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
35087053	Population structure analysis and laboratory monitoring of Shigella by core-genome multilocus sequence typing	2022	10.1038/s41467-022-28121-1	cgMLST , shared allele, NJ tree
35862685	The Population Genomics of Increased Virulence and Antibiotic Resistance in Human Commensal *Escherichia coli* over 30 Years in France	2022	10.1128/aem.00664-22	Maximum likelihood SNP tree against reference
35935201	Molecular characterization of extended spectrum cephalosporin resistant *Escherichia coli* isolated from livestock and in-contact humans in Southeast Nigeria	2022	10.3389/fmicb.2022.937968	Maximum likelihood SNP tree against reference
36695602	Characterization of *Escherichia coli* and Other Enterobacterales Resistant to Extended-Spectrum Cephalosporins Isolated from Dairy Manure in Ontario, Canada	2023	10.1128/aem.01869-22	Maximum likelihood SNP tree against reference
32826992	The population genetics of pathogenic *Escherichia coli*	2021	10.1038/s41579-020-0416-x	Roary and RaxML
33528112	Isolation and characterization of Escherichia albertii in poultry at the pre-harvest level	2021	10.1111/zph.12812	

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
33664715	Fecal Carriage and Genetic Characterization of CTX-M-1/9/1-Producing *Escherichia coli* From Healthy Humans in Hangzhou, China	2021	10.3389/fmicb.2021.616687	core genome alignment and SNP calling
33716997	Genomic Characterization of Two Shiga Toxin-Converting Bacteriophages Induced From Environmental Shiga Toxin-Producing *Escherichia coli*	2021	10.3389/fmicb.2021.587696	
35483624	Dynamics of carbapenemase-producing Enterobacterales intestinal colonisation in the elderly population after hospital discharge, Italy, 2018-2020	2022	10.1016/j.ijantimicag.2022.106594	
35847090	CARB-ES-19 Multicenter Study of Carbapenemase-Producing Klebsiella pneumoniae and *Escherichia coli* From All Spanish Provinces Reveals Interregional Spread of High-Risk Clones Such as ST307/OXA-48 and ST512/KPC-3	2022	10.3389/fmicb.2022.918362	
35989605	The effect of sequencing and assembly on the inference of horizontal gene transfer on chromosomal and plasmid phylogenies	2022	10.1098/rstb.2021.0245	

Appendix 1: Literature review of 54 studies on Pubmed (date accessed 07/08/2023), using the terms (*Escherichia coli*) AND (whole genome) AND (sequencing) AND (phylogenetic) AND (population), and subset to papers from 2020 onwards.

PMID	Title	Year	DOI	Phylogenomic method
35997596	Contrasting long-term dynamics of antimicrobial resistance and virulence plasmids in Salmonella Typhimurium from animals	2022	10.1099/mgen.0.000826	
36556442	Emergence and Dissemination of Extraintestinal Pathogenic High-Risk International Clones of *Escherichia coli*	2022	10.3390/life12122077	
36916881	Dissemination of carbapenemase-producing Enterobacterales in Ireland from 2012 to 2017: a retrospective genomic surveillance study	2023	10.1099/mgen.0.000924	VLKC, SNP sites in resulting alignments.
37143741	Prevalence and molecular characteristics of polymyxin-resistant Enterobacterales in a Chinese tertiary teaching hospital	2023	10.3389/fcimb.2023.1118122	Maximum likelihood SNP tree against reference
37272818	Psychrobacter Phage Encoding an Antibiotics Resistance Gene Represents a Novel Caudoviral Family	2023	10.1128/spectrum.05335-22	

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0001-1	PRJNA1177805	SRR31116043	SAMN44451565	117	G	2018-05	Chicken
EC18CH-0001-2	PRJNA1177805	SRR31116042	SAMN44451566	3107	A	2018-05	Chicken
EC18CH-0001-3	PRJNA1177805	SRR31115257	SAMN44451567	10	A	2018-05	Chicken
EC18CH-0001-4	PRJNA1177805	SRR31115788	SAMN44451568	191	A	2018-05	Chicken
EC18CH-0002-1	PRJNA1177805	SRR31115195	SAMN44451569	3107	A	2018-05	Chicken
EC18CH-0002-2	PRJNA1177805	SRR31115389	SAMN44451570	4038	B1	2018-05	Chicken
EC18CH-0002-3	PRJNA1177805	SRR31115646	SAMN44451571	1485	F	2018-05	Chicken
EC18CH-0002-4	PRJNA1177805	SRR31115888	SAMN44451572	1485	F	2018-05	Chicken
EC18CH-0003-1	PRJNA1177805	SRR31115873	SAMN44451573	155	B1	2018-05	Chicken
EC18CH-0003-2	PRJNA1177805	SRR31115284	SAMN44451574	155	B1	2018-05	Chicken
EC18CH-0003-3	PRJNA1177805	SRR31116041	SAMN44451575	155	B1	2018-05	Chicken
EC18CH-0003-4	PRJNA1177805	SRR31116030	SAMN44451576	155	B1	2018-05	Chicken
EC18CH-0004-1	PRJNA1177805	SRR31116019	SAMN44451577	1611	B1	2018-05	Chicken
EC18CH-0004-2	PRJNA1177805	SRR31115848	SAMN44451578	48	A	2018-05	Chicken
EC18CH-0004-3	PRJNA1177805	SRR31115837	SAMN44451579	48	A	2018-05	Chicken
EC18CH-0005-1	PRJNA1177805	SRR31115826	SAMN44451580	117	G	2018-05	Chicken
EC18CH-0005-2	PRJNA1177805	SRR31115478	SAMN44451581	69	D	2018-05	Chicken
EC18CH-0005-3	PRJNA1177805	SRR31115467	SAMN44451582	101	B1	2018-05	Chicken
EC18CH-0005-4	PRJNA1177805	SRR31115456	SAMN44451583	117	G	2018-05	Chicken
EC18CH-0006-1	PRJNA1177805	SRR31115268	SAMN44451584	1640	E	2018-05	Chicken
EC18CH-0006-2	PRJNA1177805	SRR31115256	SAMN44451585	1 Novel	A	2018-05	Chicken
EC18CH-0006-3	PRJNA1177805	SRR31115245	SAMN44451586	1485	F	2018-05	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0006-4	PRJNA1177805	SRR31116004	SAMN44451587	10	A	2018-05	Chicken
EC18CH-0007-1	PRJNA1177805	SRR31115993	SAMN44451588	58	B1	2018-05	Chicken
EC18CH-0007-2	PRJNA1177805	SRR31115982	SAMN44451589	1148	B1	2018-05	Chicken
EC18CH-0007-3	PRJNA1177805	SRR31115233	SAMN44451590	1148	B1	2018-05	Chicken
EC18CH-0007-4	PRJNA1177805	SRR31115222	SAMN44451591	58	B1	2018-05	Chicken
EC18CH-0008-1	PRJNA1177805	SRR31115211	SAMN44451592	88	C	2018-05	Chicken
EC18CH-0008-2	PRJNA1177805	SRR31115810	SAMN44451593	10	A	2018-05	Chicken
EC18CH-0008-3	PRJNA1177805	SRR31115799	SAMN44451594	1148	B1	2018-05	Chicken
EC18CH-0008-4	PRJNA1177805	SRR31115450	SAMN44451595	38	D	2018-05	Chicken
EC18CH-0009-1	PRJNA1177805	SRR31115439	SAMN44451596	140	B2	2018-06	Chicken
EC18CH-0009-2	PRJNA1177805	SRR31115428	SAMN44451597	57	E	2018-06	Chicken
EC18CH-0009-3	PRJNA1177805	SRR31115786	SAMN44451598	23	C	2018-06	Chicken
EC18CH-0010-1	PRJNA1177805	SRR31115775	SAMN44451599	1146	B1	2018-07	Chicken
EC18CH-0010-2	PRJNA1177805	SRR31115764	SAMN44451600	1844	B1	2018-07	Chicken
EC18CH-0010-3	PRJNA1177805	SRR31115977	SAMN44451601	106	D	2018-07	Chicken
EC18CH-0010-4	PRJNA1177805	SRR31115966	SAMN44451602	2001	D	2018-07	Chicken
EC18CH-0011-1	PRJNA1177805	SRR31115955	SAMN44451603	3090	B1	2018-08	Chicken
EC18CH-0011-2	PRJNA1177805	SRR31115206	SAMN44451604	10	A	2018-08	Chicken
EC18CH-0011-3	PRJNA1177805	SRR31115707	SAMN44451605	10	A	2018-08	Chicken
EC18CH-0011-4	PRJNA1177805	SRR31115696	SAMN44451606	3858	B1	2018-08	Chicken
EC18CH-0012-1	PRJNA1177805	SRR31115750	SAMN44451607	117	G	2018-08	Chicken
EC18CH-0012-2	PRJNA1177805	SRR31115739	SAMN44451608	398	A	2018-08	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0012-4	PRJNA1177805	SRR31115728	SAMN44451609	607	A	2018-08	Chicken
EC18CH-0013-1	PRJNA1177805	SRR31115941	SAMN44451610	539	B1	2018-08	Chicken
EC18CH-0013-2	PRJNA1177805	SRR31115930	SAMN44451611	539	B1	2018-08	Chicken
EC18CH-0013-4	PRJNA1177805	SRR31115919	SAMN44451612	1158	E	2018-08	Chicken
EC18CH-0014-1	PRJNA1177805	SRR31115411	SAMN44451613	1485	F	2018-08	Chicken
EC18CH-0014-2	PRJNA1177805	SRR31115400	SAMN44451614	10	A	2018-08	Chicken
EC18CH-0014-3	PRJNA1177805	SRR31115388	SAMN44451615	352	B2	2018-08	Chicken
EC18CH-0014-4	PRJNA1177805	SRR31115377	SAMN44451616	352	B2	2018-08	Chicken
EC18CH-0015-1	PRJNA1177805	SRR31115366	SAMN44451617	1056	B1	2018-08	Chicken
EC18CH-0015-2	PRJNA1177805	SRR31115355	SAMN44451618	88	C	2018-08	Chicken
EC18CH-0015-3	PRJNA1177805	SRR31115344	SAMN44451619	46	A	2018-08	Chicken
EC18CH-0016-1	PRJNA1177805	SRR31115333	SAMN44451620	117	G	2018-09	Chicken
EC18CH-0017-1	PRJNA1177805	SRR31115690	SAMN44451621	162	B1	2018-09	Chicken
EC18CH-0018-1	PRJNA1177805	SRR31115679	SAMN44451622	57	E	2018-09	Chicken
EC18CH-0019-1	PRJNA1177805	SRR31115668	SAMN44451623	1737	B1	2018-09	Chicken
EC18CH-0022-1	PRJNA1177805	SRR31115657	SAMN44451624	101	B1	2018-09	Chicken
EC18CH-0023-1	PRJNA1177805	SRR31115645	SAMN44451625	1844	B1	2018-09	Chicken
EC18CH-0025-4	PRJNA1177805	SRR31115634	SAMN44451626	57	E	2018-09	Chicken
EC18CH-0026-1	PRJNA1177805	SRR31115720	SAMN44451627	117	G	2018-09	Chicken
EC18CH-0026-3	PRJNA1177805	SRR31115709	SAMN44451628	10	A	2018-09	Chicken
EC18CH-0028-2	PRJNA1177805	SRR31115521	SAMN44451629	2520	B1	2018-09	Chicken
EC18CH-0029-2	PRJNA1177805	SRR31115622	SAMN44451630	69	D	2018-09	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0029-3	PRJNA1177805	SRR31115611	SAMN44451631	752	A	2018-09	Chicken
EC18CH-0029-4	PRJNA1177805	SRR31115600	SAMN44451632	117	G	2018-09	Chicken
EC18CH-0030-2	PRJNA1177805	SRR31115910	SAMN44451633	2 Novel	E	2018-09	Chicken
EC18CH-0030-3	PRJNA1177805	SRR31115899	SAMN44451634	1485	F	2018-09	Chicken
EC18CH-0030-4	PRJNA1177805	SRR31115887	SAMN44451635	224	B1	2018-09	Chicken
EC18CH-0033-2	PRJNA1177805	SRR31115315	SAMN44451636	1163	G	2018-09	Chicken
EC18CH-0040-1	PRJNA1177805	SRR31115304	SAMN44451637	641	B1	2018-10	Chicken
EC18CH-0040-2	PRJNA1177805	SRR31115293	SAMN44451638	155	B1	2018-10	Chicken
EC18CH-0041-1	PRJNA1177805	SRR31115506	SAMN44451639	3234	B1	2018-10	Chicken
EC18CH-0041-2	PRJNA1177805	SRR31115495	SAMN44451640	212	B1	2018-10	Chicken
EC18CH-0058-1	PRJNA1177805	SRR31115484	SAMN44451641	117	G	2018-10	Chicken
EC18CH-0058-2	PRJNA1177805	SRR31115585	SAMN44451642	4674	G	2018-10	Chicken
EC18CH-0059-1	PRJNA1177805	SRR31115574	SAMN44451643	117	G	2018-10	Chicken
EC18CH-0059-2	PRJNA1177805	SRR31115563	SAMN44451644	69	D	2018-10	Chicken
EC18CH-0066-1	PRJNA1177805	SRR31115872	SAMN44451645	155	B1	2018-10	Chicken
EC18CH-0066-2	PRJNA1177805	SRR31115861	SAMN44451646	1172	B1	2018-10	Chicken
EC18CH-0067-1	PRJNA1177805	SRR31115561	SAMN44451647	442	B1	2018-10	Chicken
EC18CH-0067-2	PRJNA1177805	SRR31115550	SAMN44451648	162	B1	2018-10	Chicken
EC18CH-0068-1	PRJNA1177805	SRR31115539	SAMN44451649	1304	B1	2018-10	Chicken
EC18CH-0068-2	PRJNA1177805	SRR31115289	SAMN44451650	3 Novel	B1	2018-10	Chicken
EC18CH-0076-1	PRJNA1177805	SRR31115288	SAMN44451651	23	C	2018-11	Chicken
EC18CH-0076-2	PRJNA1177805	SRR31115287	SAMN44451652	117	G	2018-11	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0077-1	PRJNA1177805	SRR31115286	SAMN44451653	140	B2	2018-11	Chicken
EC18CH-0077-2	PRJNA1177805	SRR31115285	SAMN44451654	135	B2	2018-11	Chicken
EC18CH-0080-1	PRJNA1177805	SRR31115283	SAMN44451655	117	G	2018-11	Chicken
EC18CH-0080-2	PRJNA1177805	SRR31115282	SAMN44451656	1172	B1	2018-11	Chicken
EC18CH-0081-1	PRJNA1177805	SRR31115281	SAMN44451657	5203	B1	2018-11	Chicken
EC18CH-0081-2	PRJNA1177805	SRR31115280	SAMN44451658	162	B1	2018-11	Chicken
EC18CH-0083-1	PRJNA1177805	SRR31115279	SAMN44451659	6286	G	2018-11	Chicken
EC18CH-0083-2	PRJNA1177805	SRR31115278	SAMN44451660	155	B1	2018-11	Chicken
EC18CH-0090-1	PRJNA1177805	SRR31115277	SAMN44451661	1276	F	2018-11	Chicken
EC18CH-0090-2	PRJNA1177805	SRR31115276	SAMN44451662	4 Novel	B1	2018-11	Chicken
EC18CH-0091-1	PRJNA1177805	SRR31115275	SAMN44451663	1485	F	2018-11	Chicken
EC18CH-0091-2	PRJNA1177805	SRR31115274	SAMN44451664	117	G	2018-11	Chicken
EC18CH-0092-1	PRJNA1177805	SRR31116040	SAMN44451665	1485	F	2018-11	Chicken
EC18CH-0092-2	PRJNA1177805	SRR31116039	SAMN44451666	744	A	2018-11	Chicken
EC18CH-0093-1	PRJNA1177805	SRR31116038	SAMN44451667	2792	E	2018-11	Chicken
EC18CH-0093-2	PRJNA1177805	SRR31116037	SAMN44451668	2792	E	2018-11	Chicken
EC18CH-0094-1	PRJNA1177805	SRR31116036	SAMN44451669	95	B2	2018-11	Chicken
EC18CH-0094-2	PRJNA1177805	SRR31116035	SAMN44451670	1485	F	2018-11	Chicken
EC18CH-0099-1	PRJNA1177805	SRR31116034	SAMN44451671	4 Novel	B1	2018-12	Chicken
EC18CH-0099-2	PRJNA1177805	SRR31116033	SAMN44451672	1775	E	2018-12	Chicken
EC18CH-0100-1	PRJNA1177805	SRR31116032	SAMN44451673	1483	B1	2018-12	Chicken
EC18CH-0100-2	PRJNA1177805	SRR31116031	SAMN44451674	1775	E	2018-12	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0103-1	PRJNA1177805	SRR31116029	SAMN44451675	117	G	2018-12	Chicken
EC18CH-0103-2	PRJNA1177805	SRR31116028	SAMN44451676	2705	A	2018-12	Chicken
EC18CH-0106-1	PRJNA1177805	SRR31116027	SAMN44451677	2772	B1	2018-12	Chicken
EC18CH-0106-2	PRJNA1177805	SRR31116026	SAMN44451678	665	A	2018-12	Chicken
EC18CH-0107-1	PRJNA1177805	SRR31116025	SAMN44451679	154	B1	2018-12	Chicken
EC18CH-0107-2	PRJNA1177805	SRR31116024	SAMN44451680	154	B1	2018-12	Chicken
EC18CH-0120-1	PRJNA1177805	SRR31116023	SAMN44451681	1011	D	2019-01	Chicken
EC18CH-0120-2	PRJNA1177805	SRR31116022	SAMN44451682	1775	E	2019-01	Chicken
EC18CH-0121-1	PRJNA1177805	SRR31116021	SAMN44451683	135	B2	2019-01	Chicken
EC18CH-0121-2	PRJNA1177805	SRR31116020	SAMN44451684	10	A	2019-01	Chicken
EC18CH-0122-1	PRJNA1177805	SRR31116018	SAMN44451685	752	A	2019-01	Chicken
EC18CH-0122-2	PRJNA1177805	SRR31116017	SAMN44451686	135	B2	2019-01	Chicken
EC18CH-0123-2	PRJNA1177805	SRR31116016	SAMN44451687	1049	B1	2019-01	Chicken
EC18CH-0124-2	PRJNA1177805	SRR31116015	SAMN44451688	2171	F	2019-01	Chicken
EC18CH-0126-1	PRJNA1177805	SRR31116014	SAMN44451689	2599	B1	2019-01	Chicken
EC18CH-0127-2	PRJNA1177805	SRR31116013	SAMN44451690	359	B1	2019-01	Chicken
EC18CH-0128-1	PRJNA1177805	SRR31116012	SAMN44451691	1286	A	2019-01	Chicken
EC18CH-0128-2	PRJNA1177805	SRR31115851	SAMN44451692	5796	G	2019-01	Chicken
EC18CH-0130-1	PRJNA1177805	SRR31115850	SAMN44451693	117	G	2019-01	Chicken
EC18CH-0130-2	PRJNA1177805	SRR31115849	SAMN44451694	4162	B1	2019-01	Chicken
EC18CH-0132-1	PRJNA1177805	SRR31115847	SAMN44451695	1844	B1	2019-01	Chicken
EC18CH-0132-2	PRJNA1177805	SRR31115846	SAMN44451696	1844	B1	2019-01	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0133-2	PRJNA1177805	SRR31115845	SAMN44451697	1148	B1	2019-01	Chicken
EC18CH-0135-2	PRJNA1177805	SRR31115844	SAMN44451698	40	B1	2019-01	Chicken
EC18CH-0137-2	PRJNA1177805	SRR31115843	SAMN44451699	1011	D	2019-02	Chicken
EC18CH-0138-1	PRJNA1177805	SRR31115842	SAMN44451700	117	G	2019-02	Chicken
EC18CH-0138-2	PRJNA1177805	SRR31115841	SAMN44451701	4993	D	2019-02	Chicken
EC18CH-0139-1	PRJNA1177805	SRR31115840	SAMN44451702	117	G	2019-02	Chicken
EC18CH-0139-2	PRJNA1177805	SRR31115839	SAMN44451703	58	B1	2019-02	Chicken
EC18CH-0140-1	PRJNA1177805	SRR31115838	SAMN44451704	10	A	2019-02	Chicken
EC18CH-0140-2	PRJNA1177805	SRR31115836	SAMN44451705	10	A	2019-02	Chicken
EC18CH-0141-2	PRJNA1177805	SRR31115835	SAMN44451706	117	G	2019-02	Chicken
EC18CH-0144-1	PRJNA1177805	SRR31115834	SAMN44451707	5 Novel	G	2019-02	Chicken
EC18CH-0144-2	PRJNA1177805	SRR31115833	SAMN44451708	297	B1	2019-02	Chicken
EC18CH-0148-1	PRJNA1177805	SRR31115832	SAMN44451709	10	A	2019-02	Chicken
EC18CH-0148-2	PRJNA1177805	SRR31115831	SAMN44451710	162	B1	2019-02	Chicken
EC18CH-0149-1	PRJNA1177805	SRR31115830	SAMN44451711	58	B1	2019-02	Chicken
EC18CH-0149-2	PRJNA1177805	SRR31115829	SAMN44451712	1286	A	2019-02	Chicken
EC18CH-0154-1	PRJNA1177805	SRR31115828	SAMN44451713	162	B1	2019-02	Chicken
EC18CH-0154-2	PRJNA1177805	SRR31115827	SAMN44451714	162	B1	2019-02	Chicken
EC18CH-0155-1	PRJNA1177805	SRR31115825	SAMN44451715	174	G	2019-02	Chicken
EC18CH-0155-2	PRJNA1177805	SRR31115824	SAMN44451716	457	F	2019-02	Chicken
EC18CH-0176-1	PRJNA1177805	SRR31115823	SAMN44451717	350	E	2019-04	Chicken
EC18CH-0176-2	PRJNA1177805	SRR31115822	SAMN44451718	355	B2	2019-04	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0192-1	PRJNA1107692	SRR28893365	SAMN41200024	6 Novel	C	2019-05	Chicken
EC18CH-0192-2	PRJNA1107692	SRR28893364	SAMN41200025	10	A	2019-05	Chicken
EC18CH-0192-3	PRJNA1107692	SRR28893221	SAMN41200026	10	A	2019-05	Chicken
EC18CH-0192-4	PRJNA1107692	SRR28893323	SAMN41200027	1485	F	2019-05	Chicken
EC18CH-0195-1	PRJNA1107692	SRR28893312	SAMN41200028	117	G	2019-05	Chicken
EC18CH-0195-2	PRJNA1107692	SRR28893301	SAMN41200029	57	E	2019-05	Chicken
EC18CH-0195-3	PRJNA1107692	SRR28893258	SAMN41200030	117	G	2019-05	Chicken
EC18CH-0195-4	PRJNA1107692	SRR28893247	SAMN41200031	7 Novel	B1	2019-05	Chicken
EC18CH-0201-1	PRJNA1107692	SRR28893236	SAMN41200032	117	G	2019-06	Chicken
EC18CH-0201-2	PRJNA1107692	SRR28893376	SAMN41200033	355	B2	2019-06	Chicken
EC18CH-0201-3	PRJNA1107692	SRR28893363	SAMN41200034	4038	B1	2019-06	Chicken
EC18CH-0201-4	PRJNA1107692	SRR28893439	SAMN41200035	3006	A	2019-06	Chicken
EC18CH-0202-1	PRJNA1107692	SRR28893428	SAMN41200036	88	C	2019-06	Chicken
EC18CH-0202-2	PRJNA1107692	SRR28893417	SAMN41200037	4162	B1	2019-06	Chicken
EC18CH-0202-3	PRJNA1107692	SRR28893351	SAMN41200038	88	C	2019-06	Chicken
EC18CH-0212-1	PRJNA1107692	SRR28893340	SAMN41200039	8 Novel	A	2019-06	Chicken
EC18CH-0212-2	PRJNA1107692	SRR28893297	SAMN41200040	8 Novel	A	2019-06	Chicken
EC18CH-0212-3	PRJNA1107692	SRR28893286	SAMN41200041	8 Novel	A	2019-06	Chicken
EC18CH-0212-4	PRJNA1107692	SRR28893275	SAMN41200042	155	B1	2019-06	Chicken
EC18CH-0215-1	PRJNA1107692	SRR28893232	SAMN41200043	354	F	2019-06	Chicken
EC18CH-0215-2	PRJNA1107692	SRR28893220	SAMN41200044	1844	B1	2019-06	Chicken
EC18CH-0215-3	PRJNA1107692	SRR28893209	SAMN41200045	354	F	2019-06	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0215-4	PRJNA1107692	SRR28893413	SAMN41200046	1146	B1	2019-06	Chicken
EC18CH-0216-1	PRJNA1107692	SRR28893402	SAMN41200047	1163	G	2019-06	Chicken
EC18CH-0216-2	PRJNA1107692	SRR28893391	SAMN41200048	101	B1	2019-06	Chicken
EC18CH-0216-3	PRJNA1107692	SRR28893328	SAMN41200049	117	G	2019-06	Chicken
EC18CH-0216-4	PRJNA1107692	SRR28893327	SAMN41200050	117	G	2019-06	Chicken
EC18CH-0222-1	PRJNA1177805	SRR31115821	SAMN44451719	162	B1	2019-06	Chicken
EC18CH-0222-2	PRJNA1177805	SRR31115820	SAMN44451720	101	B1	2019-06	Chicken
EC18CH-0223-1	PRJNA1107692	SRR28893326	SAMN41200051	1485	F	2019-06	Chicken
EC18CH-0223-2	PRJNA1107692	SRR28893325	SAMN41200052	1485	F	2019-06	Chicken
EC18CH-0223-3	PRJNA1107692	SRR28893324	SAMN41200053	1170	B2	2019-06	Chicken
EC18CH-0223-4	PRJNA1107692	SRR28893322	SAMN41200054	1485	F	2019-06	Chicken
EC18CH-0224-1	PRJNA1107692	SRR28893321	SAMN41200055	101	B1	2019-07	Chicken
EC18CH-0224-2	PRJNA1107692	SRR28893320	SAMN41200056	101	B1	2019-07	Chicken
EC18CH-0224-3	PRJNA1107692	SRR28893319	SAMN41200057	101	B1	2019-07	Chicken
EC18CH-0224-4	PRJNA1107692	SRR28893318	SAMN41200058	101	B1	2019-07	Chicken
EC18CH-0225-1	PRJNA1177805	SRR31115482	SAMN44451721	401	A	2019-07	Chicken
EC18CH-0225-2	PRJNA1177805	SRR31115481	SAMN44451722	401	A	2019-07	Chicken
EC18CH-0232-1	PRJNA1107692	SRR28893317	SAMN41200059	88	C	2019-07	Chicken
EC18CH-0232-2	PRJNA1107692	SRR28893316	SAMN41200060	117	G	2019-07	Chicken
EC18CH-0232-3	PRJNA1107692	SRR28893315	SAMN41200061	117	G	2019-07	Chicken
EC18CH-0232-4	PRJNA1107692	SRR28893314	SAMN41200062	2722	B1	2019-07	Chicken
EC18CH-0234-1	PRJNA1177805	SRR31115480	SAMN44451723	2280	B1	2019-07	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0234-2	PRJNA1177805	SRR31115479	SAMN44451724	2280	B1	2019-07	Chicken
EC18CH-0235-1	PRJNA1107692	SRR28893313	SAMN41200063	1844	B1	2019-07	Chicken
EC18CH-0235-2	PRJNA1107692	SRR28893311	SAMN41200064	1844	B1	2019-07	Chicken
EC18CH-0235-3	PRJNA1107692	SRR28893310	SAMN41200065	1844	B1	2019-07	Chicken
EC18CH-0235-4	PRJNA1107692	SRR28893309	SAMN41200066	115	D	2019-07	Chicken
EC18CH-0242-1	PRJNA1177805	SRR31115477	SAMN44451725	1140	D	2019-08	Chicken
EC18CH-0242-2	PRJNA1177805	SRR31115476	SAMN44451726	10	A	2019-08	Chicken
EC18CH-0243-1	PRJNA1107692	SRR28893308	SAMN41200067	540	A	2019-08	Chicken
EC18CH-0243-2	PRJNA1107692	SRR28893307	SAMN41200068	10	A	2019-08	Chicken
EC18CH-0243-3	PRJNA1107692	SRR28893306	SAMN41200069	1684	A	2019-08	Chicken
EC18CH-0243-4	PRJNA1107692	SRR28893305	SAMN41200070	88	C	2019-08	Chicken
EC18CH-0245-1	PRJNA1107692	SRR28893304	SAMN41200071	155	B1	2019-08	Chicken
EC18CH-0245-2	PRJNA1107692	SRR28893303	SAMN41200072	162	B1	2019-08	Chicken
EC18CH-0245-3	PRJNA1107692	SRR28893302	SAMN41200073	212	B1	2019-08	Chicken
EC18CH-0245-4	PRJNA1107692	SRR28893300	SAMN41200074	93	A	2019-08	Chicken
EC18CH-0248-1	PRJNA1177805	SRR31115475	SAMN44451727	10	A	2019-08	Chicken
EC18CH-0248-2	PRJNA1177805	SRR31115474	SAMN44451728	9 Novel	B1	2019-08	Chicken
EC18CH-0249-1	PRJNA1177805	SRR31115473	SAMN44451729	117	G	2019-08	Chicken
EC18CH-0249-2	PRJNA1177805	SRR31115472	SAMN44451730	533	B1	2019-08	Chicken
EC18CH-0251-1	PRJNA1107692	SRR28893299	SAMN41200075	124	B1	2019-08	Chicken
EC18CH-0251-2	PRJNA1107692	SRR28893298	SAMN41200076	124	B1	2019-08	Chicken
EC18CH-0251-3	PRJNA1107692	SRR28893265	SAMN41200077	57	E	2019-08	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0251-4	PRJNA1107692	SRR28893264	SAMN41200078	117	G	2019-08	Chicken
EC18CH-0252-2	PRJNA1177805	SRR31115471	SAMN44451731	117	G	2019-08	Chicken
EC18CH-0253-2	PRJNA1177805	SRR31115470	SAMN44451732	101	B1	2019-08	Chicken
EC18CH-0254-1	PRJNA1177805	SRR31115469	SAMN44451733	10	A	2019-08	Chicken
EC18CH-0254-2	PRJNA1177805	SRR31115468	SAMN44451734	156	B1	2019-08	Chicken
EC18CH-0255-1	PRJNA1177805	SRR31115466	SAMN44451735	10	A	2019-08	Chicken
EC18CH-0255-2	PRJNA1177805	SRR31115465	SAMN44451736	224	B1	2019-08	Chicken
EC18CH-0256-1	PRJNA1177805	SRR31115464	SAMN44451737	212	B1	2019-09	Chicken
EC18CH-0256-2	PRJNA1177805	SRR31115463	SAMN44451738	6635	A	2019-09	Chicken
EC18CH-0257-1	PRJNA1177805	SRR31115462	SAMN44451739	1276	F	2019-09	Chicken
EC18CH-0257-2	PRJNA1177805	SRR31115461	SAMN44451740	10	A	2019-09	Chicken
EC18CH-0258-1	PRJNA1177805	SRR31115460	SAMN44451741	88	C	2019-09	Chicken
EC18CH-0258-2	PRJNA1177805	SRR31115459	SAMN44451742	88	C	2019-09	Chicken
EC18CH-0259-2	PRJNA1177805	SRR31115458	SAMN44451743	69	D	2019-09	Chicken
EC18CH-0260-1	PRJNA1177805	SRR31115457	SAMN44451744	5375	D	2019-09	Chicken
EC18CH-0260-2	PRJNA1177805	SRR31115455	SAMN44451745	117	G	2019-09	Chicken
EC18CH-0261-1	PRJNA1177805	SRR31115454	SAMN44451746	442	B1	2019-09	Chicken
EC18CH-0261-2	PRJNA1177805	SRR31115453	SAMN44451747	155	B1	2019-09	Chicken
EC18CH-0262-1	PRJNA1177805	SRR31115452	SAMN44451748	297	B1	2019-09	Chicken
EC18CH-0262-2	PRJNA1177805	SRR31115451	SAMN44451749	297	B1	2019-09	Chicken
EC18CH-0263-1	PRJNA1177805	SRR31115273	SAMN44451750	58	B1	2019-09	Chicken
EC18CH-0263-2	PRJNA1177805	SRR31115272	SAMN44451751	602	B1	2019-09	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0264-1	PRJNA1107692	SRR28893263	SAMN41200079	351	B1	2019-09	Chicken
EC18CH-0264-2	PRJNA1107692	SRR28893262	SAMN41200080	2973	A	2019-09	Chicken
EC18CH-0264-3	PRJNA1107692	SRR28893261	SAMN41200081	1158	E	2019-09	Chicken
EC18CH-0264-4	PRJNA1107692	SRR28893260	SAMN41200082	117	G	2019-09	Chicken
EC18CH-0265-1	PRJNA1177805	SRR31115271	SAMN44451752	1011	D	2019-09	Chicken
EC18CH-0265-2	PRJNA1177805	SRR31115270	SAMN44451753	57	E	2019-09	Chicken
EC18CH-0266-1	PRJNA1107692	SRR28893259	SAMN41200083	648	F	2019-09	Chicken
EC18CH-0266-2	PRJNA1107692	SRR28893257	SAMN41200084	101	B1	2019-09	Chicken
EC18CH-0266-3	PRJNA1107692	SRR28893256	SAMN41200085	135	B2	2019-09	Chicken
EC18CH-0266-4	PRJNA1107692	SRR28893255	SAMN41200086	648	F	2019-09	Chicken
EC18CH-0267-1	PRJNA1177805	SRR31115269	SAMN44451754	10	A	2019-09	Chicken
EC18CH-0267-2	PRJNA1177805	SRR31115267	SAMN44451755	10	A	2019-09	Chicken
EC18CH-0269-1	PRJNA1177805	SRR31115266	SAMN44451756	57	E	2019-09	Chicken
EC18CH-0269-2	PRJNA1177805	SRR31115265	SAMN44451757	117	G	2019-09	Chicken
EC18CH-0270-1	PRJNA1107692	SRR28893254	SAMN41200087	23	C	2019-09	Chicken
EC18CH-0270-2	PRJNA1107692	SRR28893253	SAMN41200088	101	B1	2019-09	Chicken
EC18CH-0270-3	PRJNA1107692	SRR28893252	SAMN41200089	23	C	2019-09	Chicken
EC18CH-0270-4	PRJNA1107692	SRR28893251	SAMN41200090	23	C	2019-09	Chicken
EC18CH-0271-1	PRJNA1177805	SRR31115264	SAMN44451758	162	B1	2019-09	Chicken
EC18CH-0271-2	PRJNA1177805	SRR31115263	SAMN44451759	162	B1	2019-09	Chicken
EC18CH-0273-1	PRJNA1107692	SRR28893250	SAMN41200091	3549	B1	2019-09	Chicken
EC18CH-0273-2	PRJNA1107692	SRR28893249	SAMN41200092	57	E	2019-09	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0273-3	PRJNA1107692	SRR28893248	SAMN41200093	57	E	2019-09	Chicken
EC18CH-0273-4	PRJNA1107692	SRR28893246	SAMN41200094	57	E	2019-09	Chicken
EC18CH-0274-1	PRJNA1177805	SRR31115262	SAMN44451760	69	D	2019-09	Chicken
EC18CH-0274-2	PRJNA1177805	SRR31115261	SAMN44451761	2473	B1	2019-09	Chicken
EC18CH-0275-1	PRJNA1177805	SRR31115260	SAMN44451762	162	B1	2019-09	Chicken
EC18CH-0275-2	PRJNA1177805	SRR31115259	SAMN44451763	348	B1	2019-09	Chicken
EC18CH-0276-1	PRJNA1107692	SRR28893245	SAMN41200095	6664	D	2019-09	Chicken
EC18CH-0276-2	PRJNA1107692	SRR28893244	SAMN41200096	602	B1	2019-09	Chicken
EC18CH-0276-3	PRJNA1107692	SRR28893243	SAMN41200097	665	A	2019-09	Chicken
EC18CH-0276-4	PRJNA1107692	SRR28893242	SAMN41200098	155	B1	2019-09	Chicken
EC18CH-0279-1	PRJNA1107692	SRR28893241	SAMN41200099	1775	E	2019-09	Chicken
EC18CH-0279-2	PRJNA1107692	SRR28893240	SAMN41200100	373	A	2019-09	Chicken
EC18CH-0279-3	PRJNA1107692	SRR28893239	SAMN41200101	1551	E	2019-09	Chicken
EC18CH-0279-4	PRJNA1107692	SRR28893238	SAMN41200102	847	B1	2019-09	Chicken
EC18CH-0280-1	PRJNA1107692	SRR28893237	SAMN41200103	1049	B1	2019-10	Chicken
EC18CH-0280-2	PRJNA1107692	SRR28893235	SAMN41200104	117	G	2019-10	Chicken
EC18CH-0280-3	PRJNA1107692	SRR28893234	SAMN41200105	665	A	2019-10	Chicken
EC18CH-0281-1	PRJNA1177805	SRR31115258	SAMN44451764	677	B1	2019-10	Chicken
EC18CH-0282-1	PRJNA1107692	SRR28893384	SAMN41200106	117	G	2019-10	Chicken
EC18CH-0282-2	PRJNA1107692	SRR28893383	SAMN41200107	117	G	2019-10	Chicken
EC18CH-0282-3	PRJNA1107692	SRR28893382	SAMN41200108	117	G	2019-10	Chicken
EC18CH-0282-4	PRJNA1107692	SRR28893381	SAMN41200109	1594	A	2019-10	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0283-2	PRJNA1177805	SRR31115255	SAMN44451765	7013	A	2019-10	Chicken
EC18CH-0288-1	PRJNA1107692	SRR28893380	SAMN41200110	744	A	2019-10	Chicken
EC18CH-0288-2	PRJNA1107692	SRR28893379	SAMN41200111	2040	A	2019-10	Chicken
EC18CH-0288-3	PRJNA1107692	SRR28893378	SAMN41200112	744	A	2019-10	Chicken
EC18CH-0288-4	PRJNA1107692	SRR28893377	SAMN41200113	2599	B1	2019-10	Chicken
EC18CH-0293-1	PRJNA1107692	SRR28893375	SAMN41200114	8874	B2	2019-10	Chicken
EC18CH-0293-2	PRJNA1107692	SRR28893374	SAMN41200115	8874	B2	2019-10	Chicken
EC18CH-0293-3	PRJNA1107692	SRR28893373	SAMN41200116	8874	B2	2019-10	Chicken
EC18CH-0293-4	PRJNA1107692	SRR28893372	SAMN41200117	135	B2	2019-10	Chicken
EC18CH-0302-1	PRJNA1107692	SRR28893371	SAMN41200118	117	G	2019-11	Chicken
EC18CH-0302-2	PRJNA1107692	SRR28893370	SAMN41200119	117	G	2019-11	Chicken
EC18CH-0302-3	PRJNA1107692	SRR28893369	SAMN41200120	8611	B2	2019-11	Chicken
EC18CH-0302-4	PRJNA1107692	SRR28893368	SAMN41200121	117	G	2019-11	Chicken
EC18CH-0306-1	PRJNA1107692	SRR28893367	SAMN41200122	10	A	2019-11	Chicken
EC18CH-0306-2	PRJNA1107692	SRR28893366	SAMN41200123	10 Novel	A	2019-11	Chicken
EC18CH-0306-3	PRJNA1107692	SRR28893362	SAMN41200124	10 Novel	A	2019-11	Chicken
EC18CH-0306-4	PRJNA1107692	SRR28893448	SAMN41200125	206	A	2019-11	Chicken
EC18CH-0310-1	PRJNA1107692	SRR28893447	SAMN41200126	8611	B2	2019-11	Chicken
EC18CH-0310-2	PRJNA1107692	SRR28893446	SAMN41200127	57	E	2019-11	Chicken
EC18CH-0310-3	PRJNA1107692	SRR28893445	SAMN41200128	131	B2	2019-11	Chicken
EC18CH-0311-1	PRJNA1107692	SRR28893444	SAMN41200129	48	A	2019-11	Chicken

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18CH-0311-2	PRJNA1107692	SRR28893443	SAMN41200130	58	B1	2019-11	Chicken
EC18CH-0311-3	PRJNA1107692	SRR28893442	SAMN41200131	351	B1	2019-11	Chicken
EC18CH-0311-4	PRJNA1107692	SRR28893441	SAMN41200132	641	B1	2019-11	Chicken
EC18LG-0001-1	PRJNA1177805	SRR31115254	SAMN44451766	6422	B1	2018-05	Leafy_Greens
EC18LG-0001-2	PRJNA1177805	SRR31115253	SAMN44451767	1723	B1	2018-05	Leafy_Greens
EC18LG-0001-3	PRJNA1177805	SRR31115252	SAMN44451768	212	B1	2018-05	Leafy_Greens
EC18LG-0001-4	PRJNA1177805	SRR31115251	SAMN44451769	6422	B1	2018-05	Leafy_Greens
EC18LG-0003-1	PRJNA1177805	SRR31115250	SAMN44451770	452	B2	2018-05	Leafy_Greens
EC18LG-0003-2	PRJNA1177805	SRR31115249	SAMN44451771	452	B2	2018-05	Leafy_Greens
EC18LG-0003-3	PRJNA1177805	SRR31115248	SAMN44451772	452	B2	2018-05	Leafy_Greens
EC18LG-0003-4	PRJNA1177805	SRR31115247	SAMN44451773	452	B2	2018-05	Leafy_Greens
EC18LG-0005-1	PRJNA1177805	SRR31115246	SAMN44451774	10	A	2018-05	Leafy_Greens
EC18LG-0005-2	PRJNA1177805	SRR31115244	SAMN44451775	10	A	2018-05	Leafy_Greens
EC18LG-0010-1	PRJNA1177805	SRR31115243	SAMN44451776	1421	A	2018-07	Leafy_Greens
EC18LG-0010-2	PRJNA1177805	SRR31115242	SAMN44451777	1421	A	2018-07	Leafy_Greens
EC18LG-0010-3	PRJNA1177805	SRR31116011	SAMN44451778	642	B1	2018-07	Leafy_Greens
EC18LG-0010-4	PRJNA1177805	SRR31116010	SAMN44451779	1421	A	2018-07	Leafy_Greens
EC18LG-0011-1	PRJNA1177805	SRR31116009	SAMN44451780	939	A	2018-08	Leafy_Greens
EC18LG-0011-2	PRJNA1177805	SRR31116008	SAMN44451781	939	A	2018-08	Leafy_Greens
EC18LG-0011-3	PRJNA1177805	SRR31116007	SAMN44451782	939	A	2018-08	Leafy_Greens
EC18LG-0011-4	PRJNA1177805	SRR31116006	SAMN44451783	939	A	2018-08	Leafy_Greens
EC18LG-0013-1	PRJNA1177805	SRR31116005	SAMN44451784	937	B1	2018-08	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0013-2	PRJNA1177805	SRR31116003	SAMN44451785	937	B1	2018-08	Leafy_Greens
EC18LG-0013-3	PRJNA1177805	SRR31116002	SAMN44451786	937	B1	2018-08	Leafy_Greens
EC18LG-0013-4	PRJNA1177805	SRR31116001	SAMN44451787	937	B1	2018-08	Leafy_Greens
EC18LG-0014-1	PRJNA1177805	SRR31116000	SAMN44451788	906	B1	2018-08	Leafy_Greens
EC18LG-0014-2	PRJNA1177805	SRR31115999	SAMN44451789	906	B1	2018-08	Leafy_Greens
EC18LG-0014-3	PRJNA1177805	SRR31115998	SAMN44451790	906	B1	2018-08	Leafy_Greens
EC18LG-0014-4	PRJNA1177805	SRR31115997	SAMN44451791	906	B1	2018-08	Leafy_Greens
EC18LG-0016-1	PRJNA1177805	SRR31115996	SAMN44451792	2144	B1	2018-09	Leafy_Greens
EC18LG-0016-2	PRJNA1177805	SRR31115995	SAMN44451793	69	D	2018-09	Leafy_Greens
EC18LG-0016-3	PRJNA1177805	SRR31115994	SAMN44451794	349	D	2018-09	Leafy_Greens
EC18LG-0016-4	PRJNA1177805	SRR31115992	SAMN44451795	349	D	2018-09	Leafy_Greens
EC18LG-0017-1	PRJNA1177805	SRR31115991	SAMN44451796	11 Novel	D	2018-09	Leafy_Greens
EC18LG-0017-2	PRJNA1177805	SRR31115990	SAMN44451797	11 Novel	D	2018-09	Leafy_Greens
EC18LG-0017-3	PRJNA1177805	SRR31115989	SAMN44451798	11 Novel	D	2018-09	Leafy_Greens
EC18LG-0017-4	PRJNA1177805	SRR31115988	SAMN44451799	11 Novel	D	2018-09	Leafy_Greens
EC18LG-0018-1	PRJNA1177805	SRR31115987	SAMN44451800	1148	B1	2018-09	Leafy_Greens
EC18LG-0018-2	PRJNA1177805	SRR31115986	SAMN44451801	1148	B1	2018-09	Leafy_Greens
EC18LG-0018-3	PRJNA1177805	SRR31115985	SAMN44451802	1148	B1	2018-09	Leafy_Greens
EC18LG-0018-4	PRJNA1177805	SRR31115984	SAMN44451803	1148	B1	2018-09	Leafy_Greens
EC18LG-0019-1	PRJNA1177805	SRR31115983	SAMN44451804	1049	B1	2018-09	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0020-1	PRJNA1177805	SRR31115981	SAMN44451805	906	B1	2018-09	Leafy_Greens
EC18LG-0020-2	PRJNA1177805	SRR31115980	SAMN44451806	906	B1	2018-09	Leafy_Greens
EC18LG-0020-3	PRJNA1177805	SRR31115241	SAMN44451807	906	B1	2018-09	Leafy_Greens
EC18LG-0020-4	PRJNA1177805	SRR31115240	SAMN44451808	12 Novel	B1	2018-09	Leafy_Greens
EC18LG-0021-1	PRJNA1177805	SRR31115239	SAMN44451809	1730	B1	2018-09	Leafy_Greens
EC18LG-0021-2	PRJNA1177805	SRR31115238	SAMN44451810	4038	B1	2018-09	Leafy_Greens
EC18LG-0021-3	PRJNA1177805	SRR31115237	SAMN44451811	13 Novel	B1	2018-09	Leafy_Greens
EC18LG-0021-4	PRJNA1177805	SRR31115236	SAMN44451812	2005	B1	2018-09	Leafy_Greens
EC18LG-0022-1	PRJNA1177805	SRR31115235	SAMN44451813	2602	B1	2018-09	Leafy_Greens
EC18LG-0024-1	PRJNA1177805	SRR31115234	SAMN44451814	219	E	2018-09	Leafy_Greens
EC18LG-0024-2	PRJNA1177805	SRR31115232	SAMN44451815	219	E	2018-09	Leafy_Greens
EC18LG-0024-3	PRJNA1177805	SRR31115231	SAMN44451816	1858	B2	2018-09	Leafy_Greens
EC18LG-0024-4	PRJNA1177805	SRR31115230	SAMN44451817	219	E	2018-09	Leafy_Greens
EC18LG-0026-1	PRJNA1177805	SRR31115229	SAMN44451818	4197	E	2018-09	Leafy_Greens
EC18LG-0026-2	PRJNA1177805	SRR31115228	SAMN44451819	14 Novel	A	2018-09	Leafy_Greens
EC18LG-0026-3	PRJNA1177805	SRR31115227	SAMN44451820	1049	B1	2018-09	Leafy_Greens
EC18LG-0026-4	PRJNA1177805	SRR31115226	SAMN44451821	4197	E	2018-09	Leafy_Greens
EC18LG-0028-1	PRJNA1177805	SRR31115225	SAMN44451822	2161	B1	2018-09	Leafy_Greens
EC18LG-0028-2	PRJNA1177805	SRR31115224	SAMN44451823	2161	B1	2018-09	Leafy_Greens
EC18LG-0028-3	PRJNA1177805	SRR31115223	SAMN44451824	2161	B1	2018-09	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0028-4	PRJNA1177805	SRR31115221	SAMN44451825	2161	B1	2018-09	Leafy_Greens
EC18LG-0029-1	PRJNA1177805	SRR31115220	SAMN44451826	641	B1	2018-09	Leafy_Greens
EC18LG-0029-2	PRJNA1177805	SRR31115219	SAMN44451827	641	B1	2018-09	Leafy_Greens
EC18LG-0029-3	PRJNA1177805	SRR31115218	SAMN44451828	641	B1	2018-09	Leafy_Greens
EC18LG-0029-4	PRJNA1177805	SRR31115217	SAMN44451829	641	B1	2018-09	Leafy_Greens
EC18LG-0030-1	PRJNA1177805	SRR31115216	SAMN44451830	3167	B1	2018-09	Leafy_Greens
EC18LG-0030-2	PRJNA1177805	SRR31115215	SAMN44451831	3167	B1	2018-09	Leafy_Greens
EC18LG-0030-3	PRJNA1177805	SRR31115214	SAMN44451832	69	D	2018-09	Leafy_Greens
EC18LG-0030-4	PRJNA1177805	SRR31115213	SAMN44451833	69	D	2018-09	Leafy_Greens
EC18LG-0032-1	PRJNA1177805	SRR31115212	SAMN44451834	15 Novel	B1	2018-09	Leafy_Greens
EC18LG-0032-2	PRJNA1177805	SRR31115210	SAMN44451835	15 Novel	B1	2018-09	Leafy_Greens
EC18LG-0032-3	PRJNA1177805	SRR31115819	SAMN44451836	15 Novel	B1	2018-09	Leafy_Greens
EC18LG-0032-4	PRJNA1177805	SRR31115818	SAMN44451837	15 Novel	B1	2018-09	Leafy_Greens
EC18LG-0037-1	PRJNA1177805	SRR31115817	SAMN44451838	4118	D	2018-09	Leafy_Greens
EC18LG-0037-2	PRJNA1177805	SRR31115816	SAMN44451839	4118	D	2018-09	Leafy_Greens
EC18LG-0037-3	PRJNA1177805	SRR31115815	SAMN44451840	4118	D	2018-09	Leafy_Greens
EC18LG-0037-4	PRJNA1177805	SRR31115814	SAMN44451841	4118	D	2018-09	Leafy_Greens
EC18LG-0039-2	PRJNA1177805	SRR31115813	SAMN44451842	16 Novel	A	2018-09	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0039-4	PRJNA1177805	SRR31115812	SAMN44451843	16 Novel	A	2018-09	Leafy_Greens
EC18LG-0041-1	PRJNA1177805	SRR31115811	SAMN44451844	108	D	2018-10	Leafy_Greens
EC18LG-0041-2	PRJNA1177805	SRR31115809	SAMN44451845	108	D	2018-10	Leafy_Greens
EC18LG-0041-3	PRJNA1177805	SRR31115808	SAMN44451846	108	D	2018-10	Leafy_Greens
EC18LG-0041-4	PRJNA1177805	SRR31115807	SAMN44451847	108	D	2018-10	Leafy_Greens
EC18LG-0043-1	PRJNA1177805	SRR31115806	SAMN44451848	1629	E	2018-10	Leafy_Greens
EC18LG-0043-2	PRJNA1177805	SRR31115805	SAMN44451849	1629	E	2018-10	Leafy_Greens
EC18LG-0043-3	PRJNA1177805	SRR31115804	SAMN44451850	1629	E	2018-10	Leafy_Greens
EC18LG-0043-4	PRJNA1177805	SRR31115803	SAMN44451851	1629	E	2018-10	Leafy_Greens
EC18LG-0044-1	PRJNA1177805	SRR31115802	SAMN44451852	155	B1	2018-10	Leafy_Greens
EC18LG-0044-2	PRJNA1177805	SRR31115801	SAMN44451853	155	B1	2018-10	Leafy_Greens
EC18LG-0044-3	PRJNA1177805	SRR31115800	SAMN44451854	155	B1	2018-10	Leafy_Greens
EC18LG-0044-4	PRJNA1177805	SRR31115798	SAMN44451855	155	B1	2018-10	Leafy_Greens
EC18LG-0045-1	PRJNA1177805	SRR31115797	SAMN44451856	17 Novel	A	2018-10	Leafy_Greens
EC18LG-0045-2	PRJNA1177805	SRR31115796	SAMN44451857	17 Novel	A	2018-10	Leafy_Greens
EC18LG-0045-3	PRJNA1177805	SRR31115795	SAMN44451858	17 Novel	A	2018-10	Leafy_Greens
EC18LG-0045-4	PRJNA1177805	SRR31115794	SAMN44451859	17 Novel	A	2018-10	Leafy_Greens
EC18LG-0071-1	PRJNA1177805	SRR31115793	SAMN44451860	69	D	2018-10	Leafy_Greens
EC18LG-0071-2	PRJNA1177805	SRR31115792	SAMN44451861	69	D	2018-10	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0096-1	PRJNA1177805	SRR31115791	SAMN44451862	131	B2	2018-12	Leafy_Greens
EC18LG-0097-1	PRJNA1177805	SRR31115790	SAMN44451863	2973	A	2018-12	Leafy_Greens
EC18LG-0097-2	PRJNA1177805	SRR31115789	SAMN44451864	130	E	2018-12	Leafy_Greens
EC18LG-0100-1	PRJNA1177805	SRR31115449	SAMN44451865	5765	B1	2018-12	Leafy_Greens
EC18LG-0102-1	PRJNA1177805	SRR31115448	SAMN44451866	602	B1	2018-12	Leafy_Greens
EC18LG-0102-2	PRJNA1177805	SRR31115447	SAMN44451867	602	B1	2018-12	Leafy_Greens
EC18LG-0203-1	PRJNA1107692	SRR28893440	SAMN41200133	23	C	2019-06	Leafy_Greens
EC18LG-0207-1	PRJNA1107692	SRR28893438	SAMN41200134	111	B1	2019-06	Leafy_Greens
EC18LG-0207-2-R	PRJNA1177805	SRR31115446	SAMN44451868	111	B1	2019-06	Leafy_Greens
EC18LG-0207-3-R	PRJNA1177805	SRR31115445	SAMN44451869	1148	B1	2019-06	Leafy_Greens
EC18LG-0207-4-R	PRJNA1177805	SRR31115444	SAMN44451870	111	B1	2019-06	Leafy_Greens
EC18LG-0208-1-R	PRJNA1177805	SRR31115443	SAMN44451871	191	A	2019-06	Leafy_Greens
EC18LG-0208-2-R	PRJNA1177805	SRR31115442	SAMN44451872	191	A	2019-06	Leafy_Greens
EC18LG-0208-3-R	PRJNA1177805	SRR31115441	SAMN44451873	10176	B1	2019-06	Leafy_Greens
EC18LG-0211-1	PRJNA1107692	SRR28893431	SAMN41200141	2522	B1	2019-06	Leafy_Greens
EC18LG-0211-2	PRJNA1107692	SRR28893430	SAMN41200142	2522	B1	2019-06	Leafy_Greens
EC18LG-0211-3	PRJNA1107692	SRR28893429	SAMN41200143	6603	B2	2019-06	Leafy_Greens
EC18LG-0211-4	PRJNA1107692	SRR28893427	SAMN41200144	2522	B1	2019-06	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0216-1	PRJNA1107692	SRR28893426	SAMN41200145	3576	B1	2019-06	Leafy_Greens
EC18LG-0216-2	PRJNA1107692	SRR28893425	SAMN41200146	3576	B1	2019-06	Leafy_Greens
EC18LG-0216-3	PRJNA1107692	SRR28893424	SAMN41200147	3576	B1	2019-06	Leafy_Greens
EC18LG-0218-1	PRJNA1107692	SRR28893423	SAMN41200148	10	A	2019-06	Leafy_Greens
EC18LG-0218-2	PRJNA1107692	SRR28893422	SAMN41200149	2973	A	2019-06	Leafy_Greens
EC18LG-0218-3	PRJNA1107692	SRR28893421	SAMN41200150	154	B1	2019-06	Leafy_Greens
EC18LG-0218-4	PRJNA1107692	SRR28893420	SAMN41200151	57	E	2019-06	Leafy_Greens
EC18LG-0222-1	PRJNA1107692	SRR28893419	SAMN41200152	5478	B1	2019-06	Leafy_Greens
EC18LG-0222-2	PRJNA1107692	SRR28893418	SAMN41200153	3568	cladeIII	2019-06	Leafy_Greens
EC18LG-0222-3	PRJNA1107692	SRR28893361	SAMN41200154	5478	B1	2019-06	Leafy_Greens
EC18LG-0233-1	PRJNA1107692	SRR28893360	SAMN41200155	1684	A	2019-07	Leafy_Greens
EC18LG-0233-2	PRJNA1107692	SRR28893359	SAMN41200156	5614	B1	2019-07	Leafy_Greens
EC18LG-0233-3	PRJNA1107692	SRR28893358	SAMN41200157	5156	F	2019-07	Leafy_Greens
EC18LG-0239-1	PRJNA1107692	SRR28893357	SAMN41200158	642	B1	2019-07	Leafy_Greens
EC18LG-0239-2	PRJNA1107692	SRR28893356	SAMN41200159	642	B1	2019-07	Leafy_Greens
EC18LG-0239-3	PRJNA1107692	SRR28893355	SAMN41200160	642	B1	2019-07	Leafy_Greens
EC18LG-0239-4	PRJNA1107692	SRR28893354	SAMN41200161	642	B1	2019-07	Leafy_Greens
EC18LG-0247-1	PRJNA1107692	SRR28893353	SAMN41200162	1167	B1	2019-08	Leafy_Greens
EC18LG-0250-1	PRJNA1177805	SRR31115440	SAMN44451874	4038	B1	2019-08	Leafy_Greens
EC18LG-0250-2	PRJNA1177805	SRR31115438	SAMN44451875	4038	B1	2019-08	Leafy_Greens
EC18LG-0254-1	PRJNA1107692	SRR28893352	SAMN41200163	18 Novel	E	2019-08	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0254-2	PRJNA1107692	SRR28893350	SAMN41200164	18 Novel	E	2019-08	Leafy_Greens
EC18LG-0254-3	PRJNA1107692	SRR28893349	SAMN41200165	18 Novel	E	2019-08	Leafy_Greens
EC18LG-0265-1	PRJNA1107692	SRR28893348	SAMN41200166	6188	B1	2019-09	Leafy_Greens
EC18LG-0265-2	PRJNA1107692	SRR28893347	SAMN41200167	196	D	2019-09	Leafy_Greens
EC18LG-0265-3	PRJNA1107692	SRR28893346	SAMN41200168	6188	B1	2019-09	Leafy_Greens
EC18LG-0265-4	PRJNA1107692	SRR28893345	SAMN41200169	28	B2	2019-09	Leafy_Greens
EC18LG-0271-1	PRJNA1177805	SRR31115437	SAMN44451876	297	B1	2019-09	Leafy_Greens
EC18LG-0271-2	PRJNA1177805	SRR31115436	SAMN44451877	297	B1	2019-09	Leafy_Greens
EC18LG-0274-1	PRJNA1107692	SRR28893344	SAMN41200170	19 Novel	B1	2019-09	Leafy_Greens
EC18LG-0274-2	PRJNA1107692	SRR28893343	SAMN41200171	19 Novel	B1	2019-09	Leafy_Greens
EC18LG-0274-3	PRJNA1107692	SRR28893342	SAMN41200172	19 Novel	B1	2019-09	Leafy_Greens
EC18LG-0274-4	PRJNA1107692	SRR28893341	SAMN41200173	19 Novel	B1	2019-09	Leafy_Greens
EC18LG-0277-1	PRJNA1107692	SRR28893339	SAMN41200174	1844	B1	2019-09	Leafy_Greens
EC18LG-0283-1	PRJNA1107692	SRR28893338	SAMN41200175	154	B1	2019-10	Leafy_Greens
EC18LG-0283-2	PRJNA1107692	SRR28893337	SAMN41200176	154	B1	2019-10	Leafy_Greens
EC18LG-0283-3	PRJNA1107692	SRR28893336	SAMN41200177	154	B1	2019-10	Leafy_Greens
EC18LG-0283-4	PRJNA1107692	SRR28893335	SAMN41200178	154	B1	2019-10	Leafy_Greens
EC18LG-0287-1	PRJNA1107692	SRR28893334	SAMN41200179	56	B1	2019-10	Leafy_Greens

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18LG-0287-2	PRJNA1107692	SRR28893333	SAMN41200180	56	B1	2019-10	Leafy_Greens
EC18LG-0287-3	PRJNA1107692	SRR28893332	SAMN41200181	56	B1	2019-10	Leafy_Greens
EC18LG-0297-1	PRJNA1107692	SRR28893331	SAMN41200182	196	B1	2019-11	Leafy_Greens
EC18LG-0297-2	PRJNA1107692	SRR28893330	SAMN41200183	34	A	2019-11	Leafy_Greens
EC18LG-0297-3	PRJNA1107692	SRR28893296	SAMN41200184	20 Novel	B1	2019-11	Leafy_Greens
EC18LG-0298-1	PRJNA1107692	SRR28893295	SAMN41200185	448	B1	2019-11	Leafy_Greens
EC18LG-0298-2	PRJNA1107692	SRR28893294	SAMN41200186	448	B1	2019-11	Leafy_Greens
EC18LG-0302-1	PRJNA1177805	SRR31115435	SAMN44451878	5328	B2	2019-11	Leafy_Greens
EC18LG-0302-2	PRJNA1177805	SRR31115434	SAMN44451879	5328	B2	2019-11	Leafy_Greens
EC18PK-0002-1	PRJNA1177805	SRR31115433	SAMN44451880	625	B2	2018-05	Pork
EC18PK-0002-2	PRJNA1177805	SRR31115432	SAMN44451881	710	A	2018-05	Pork
EC18PK-0002-3	PRJNA1177805	SRR31115431	SAMN44451882	1434	A	2018-05	Pork
EC18PK-0002-4	PRJNA1177805	SRR31115430	SAMN44451883	1434	A	2018-05	Pork
EC18PK-0003-1	PRJNA1177805	SRR31115429	SAMN44451884	10	A	2018-05	Pork
EC18PK-0003-2	PRJNA1177805	SRR31115427	SAMN44451885	10	A	2018-05	Pork
EC18PK-0003-3	PRJNA1177805	SRR31115426	SAMN44451886	21 Novel	A	2018-05	Pork
EC18PK-0003-4	PRJNA1177805	SRR31115425	SAMN44451887	10	A	2018-05	Pork
EC18PK-0007-1	PRJNA1177805	SRR31115424	SAMN44451888	345	B1	2018-05	Pork
EC18PK-0007-2	PRJNA1177805	SRR31115423	SAMN44451889	345	B1	2018-05	Pork
EC18PK-0007-3	PRJNA1177805	SRR31115422	SAMN44451890	23	C	2018-05	Pork
EC18PK-0007-4	PRJNA1177805	SRR31115421	SAMN44451891	23	C	2018-05	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0008-1	PRJNA1177805	SRR31115420	SAMN44451892	1011	D	2018-05	Pork
EC18PK-0008-2	PRJNA1177805	SRR31115419	SAMN44451893	155	B1	2018-05	Pork
EC18PK-0008-3	PRJNA1177805	SRR31115787	SAMN44451894	1148	B1	2018-05	Pork
EC18PK-0008-4	PRJNA1177805	SRR31115785	SAMN44451895	155	B1	2018-05	Pork
EC18PK-0010-1	PRJNA1177805	SRR31115784	SAMN44451896	942	B1	2018-07	Pork
EC18PK-0010-2	PRJNA1177805	SRR31115783	SAMN44451897	1434	A	2018-07	Pork
EC18PK-0010-3	PRJNA1177805	SRR31115782	SAMN44451898	1434	A	2018-07	Pork
EC18PK-0010-4	PRJNA1177805	SRR31115781	SAMN44451899	1434	A	2018-07	Pork
EC18PK-0011-1	PRJNA1177805	SRR31115780	SAMN44451900	2531	A	2018-08	Pork
EC18PK-0011-2	PRJNA1177805	SRR31115779	SAMN44451901	2531	A	2018-08	Pork
EC18PK-0011-3	PRJNA1177805	SRR31115778	SAMN44451902	2531	A	2018-08	Pork
EC18PK-0011-4	PRJNA1177805	SRR31115777	SAMN44451903	2531	A	2018-08	Pork
EC18PK-0012-1	PRJNA1177805	SRR31115776	SAMN44451904	1429	A	2018-08	Pork
EC18PK-0012-2	PRJNA1177805	SRR31115774	SAMN44451905	1429	A	2018-08	Pork
EC18PK-0012-3	PRJNA1177805	SRR31115773	SAMN44451906	1429	A	2018-08	Pork
EC18PK-0013-1	PRJNA1177805	SRR31115772	SAMN44451907	10	A	2018-08	Pork
EC18PK-0013-2	PRJNA1177805	SRR31115771	SAMN44451908	1972	A	2018-08	Pork
EC18PK-0013-3	PRJNA1177805	SRR31115770	SAMN44451909	1972	A	2018-08	Pork
EC18PK-0013-4	PRJNA1177805	SRR31115769	SAMN44451910	216	A	2018-08	Pork
EC18PK-0014-1	PRJNA1177805	SRR31115768	SAMN44451911	4704	A	2018-08	Pork
EC18PK-0014-2	PRJNA1177805	SRR31115767	SAMN44451912	4704	A	2018-08	Pork
EC18PK-0014-3	PRJNA1177805	SRR31115766	SAMN44451913	4704	A	2018-08	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0014-4	PRJNA1177805	SRR31115765	SAMN44451914	4704	A	2018-08	Pork
EC18PK-0015-2	PRJNA1177805	SRR31115763	SAMN44451915	111	B1	2018-08	Pork
EC18PK-0015-3	PRJNA1177805	SRR31115762	SAMN44451916	124	B1	2018-08	Pork
EC18PK-0015-4	PRJNA1177805	SRR31115761	SAMN44451917	111	B1	2018-08	Pork
EC18PK-0016-1	PRJNA1177805	SRR31115760	SAMN44451918	216	A	2018-09	Pork
EC18PK-0016-2	PRJNA1177805	SRR31115759	SAMN44451919	216	A	2018-09	Pork
EC18PK-0016-3	PRJNA1177805	SRR31115758	SAMN44451920	216	A	2018-09	Pork
EC18PK-0016-4	PRJNA1177805	SRR31115757	SAMN44451921	216	A	2018-09	Pork
EC18PK-0017-1	PRJNA1177805	SRR31115756	SAMN44451922	101	B1	2018-09	Pork
EC18PK-0017-2	PRJNA1177805	SRR31115979	SAMN44451923	101	B1	2018-09	Pork
EC18PK-0017-3	PRJNA1177805	SRR31115978	SAMN44451924	101	B1	2018-09	Pork
EC18PK-0017-4	PRJNA1177805	SRR31115976	SAMN44451925	101	B1	2018-09	Pork
EC18PK-0018-1	PRJNA1177805	SRR31115975	SAMN44451926	541	A	2018-09	Pork
EC18PK-0018-4	PRJNA1177805	SRR31115974	SAMN44451927	131	B2	2018-09	Pork
EC18PK-0019-1	PRJNA1177805	SRR31115973	SAMN44451928	1718	A	2018-09	Pork
EC18PK-0019-3	PRJNA1177805	SRR31115972	SAMN44451929	1718	A	2018-09	Pork
EC18PK-0019-4	PRJNA1177805	SRR31115971	SAMN44451930	1718	A	2018-09	Pork
EC18PK-0020-1	PRJNA1177805	SRR31115970	SAMN44451931	399	A	2018-09	Pork
EC18PK-0021-1	PRJNA1177805	SRR31115969	SAMN44451932	5474	B1	2018-09	Pork
EC18PK-0021-2	PRJNA1177805	SRR31115968	SAMN44451933	5474	B1	2018-09	Pork
EC18PK-0021-3	PRJNA1177805	SRR31115967	SAMN44451934	5474	B1	2018-09	Pork
EC18PK-0021-4	PRJNA1177805	SRR31115965	SAMN44451935	5474	B1	2018-09	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0022-1	PRJNA1177805	SRR31115964	SAMN44451936	101	B1	2018-09	Pork
EC18PK-0022-2	PRJNA1177805	SRR31115963	SAMN44451937	13	B1	2018-09	Pork
EC18PK-0022-3	PRJNA1177805	SRR31115962	SAMN44451938	10	A	2018-09	Pork
EC18PK-0022-4	PRJNA1177805	SRR31115961	SAMN44451939	88	C	2018-09	Pork
EC18PK-0023-1	PRJNA1177805	SRR31115960	SAMN44451940	101	B1	2018-09	Pork
EC18PK-0024-1	PRJNA1177805	SRR31115959	SAMN44451941	7940	E	2018-09	Pork
EC18PK-0024-2	PRJNA1177805	SRR31115958	SAMN44451942	7940	E	2018-09	Pork
EC18PK-0024-3	PRJNA1177805	SRR31115957	SAMN44451943	7940	E	2018-09	Pork
EC18PK-0024-4	PRJNA1177805	SRR31115956	SAMN44451944	7940	E	2018-09	Pork
EC18PK-0025-1	PRJNA1177805	SRR31115954	SAMN44451945	206	A	2018-09	Pork
EC18PK-0025-2	PRJNA1177805	SRR31115953	SAMN44451946	10	A	2018-09	Pork
EC18PK-0025-3	PRJNA1177805	SRR31115952	SAMN44451947	155	B1	2018-09	Pork
EC18PK-0025-4	PRJNA1177805	SRR31115951	SAMN44451948	10	A	2018-09	Pork
EC18PK-0026-3	PRJNA1177805	SRR31115950	SAMN44451949	2531	A	2018-09	Pork
EC18PK-0027-1	PRJNA1177805	SRR31115949	SAMN44451950	2144	B1	2018-09	Pork
EC18PK-0027-2	PRJNA1177805	SRR31115948	SAMN44451951	22 Novel	G	2018-09	Pork
EC18PK-0027-3	PRJNA1177805	SRR31115209	SAMN44451952	602	B1	2018-09	Pork
EC18PK-0027-4	PRJNA1177805	SRR31115208	SAMN44451953	602	B1	2018-09	Pork
EC18PK-0028-1	PRJNA1177805	SRR31115207	SAMN44451954	677	B1	2018-09	Pork
EC18PK-0028-2	PRJNA1177805	SRR31115205	SAMN44451955	7940	E	2018-09	Pork
EC18PK-0028-3	PRJNA1177805	SRR31115204	SAMN44451956	677	B1	2018-09	Pork
EC18PK-0028-4	PRJNA1177805	SRR31115203	SAMN44451957	677	B1	2018-09	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0029-1	PRJNA1177805	SRR31115202	SAMN44451958	677	B1	2018-09	Pork
EC18PK-0029-2	PRJNA1177805	SRR31115201	SAMN44451959	677	B1	2018-09	Pork
EC18PK-0029-3	PRJNA1177805	SRR31115200	SAMN44451960	677	B1	2018-09	Pork
EC18PK-0029-4	PRJNA1177805	SRR31115199	SAMN44451961	635	A	2018-09	Pork
EC18PK-0030-1	PRJNA1177805	SRR31115198	SAMN44451962	1434	A	2018-09	Pork
EC18PK-0030-2	PRJNA1177805	SRR31115197	SAMN44451963	23 Novel	A	2018-09	Pork
EC18PK-0030-3	PRJNA1177805	SRR31115196	SAMN44451964	23 Novel	A	2018-09	Pork
EC18PK-0030-4	PRJNA1177805	SRR31115706	SAMN44451965	1434	A	2018-09	Pork
EC18PK-0032-1	PRJNA1177805	SRR31115705	SAMN44451966	10	A	2018-09	Pork
EC18PK-0032-2	PRJNA1177805	SRR31115704	SAMN44451967	10	A	2018-09	Pork
EC18PK-0032-3	PRJNA1177805	SRR31115703	SAMN44451968	10	A	2018-09	Pork
EC18PK-0032-4	PRJNA1177805	SRR31115702	SAMN44451969	10	A	2018-09	Pork
EC18PK-0033-1	PRJNA1177805	SRR31115701	SAMN44451970	201	B1	2018-09	Pork
EC18PK-0033-2	PRJNA1177805	SRR31115700	SAMN44451971	1248	B1	2018-09	Pork
EC18PK-0033-3	PRJNA1177805	SRR31115699	SAMN44451972	1079	B1	2018-09	Pork
EC18PK-0033-4	PRJNA1177805	SRR31115698	SAMN44451973	1079	B1	2018-09	Pork
EC18PK-0035-1	PRJNA1177805	SRR31115697	SAMN44451974	5082	B1	2018-09	Pork
EC18PK-0035-2	PRJNA1177805	SRR31115695	SAMN44451975	2178	B1	2018-09	Pork
EC18PK-0035-3	PRJNA1177805	SRR31115694	SAMN44451976	648	F	2018-09	Pork
EC18PK-0035-4	PRJNA1177805	SRR31115693	SAMN44451977	5082	B1	2018-09	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0036-1	PRJNA1177805	SRR31115692	SAMN44451978	24 Novel	A	2018-09	Pork
EC18PK-0036-2	PRJNA1177805	SRR31115691	SAMN44451979	635	A	2018-09	Pork
EC18PK-0036-3	PRJNA1177805	SRR31115755	SAMN44451980	24 Novel	A	2018-09	Pork
EC18PK-0036-4	PRJNA1177805	SRR31115754	SAMN44451981	24 Novel	A	2018-09	Pork
EC18PK-0038-1	PRJNA1177805	SRR31115753	SAMN44451982	641	B1	2018-09	Pork
EC18PK-0038-2	PRJNA1177805	SRR31115752	SAMN44451983	641	B1	2018-09	Pork
EC18PK-0038-3	PRJNA1177805	SRR31115751	SAMN44451984	641	B1	2018-09	Pork
EC18PK-0038-4	PRJNA1177805	SRR31115749	SAMN44451985	641	B1	2018-09	Pork
EC18PK-0040-1	PRJNA1177805	SRR31115748	SAMN44451986	1684	A	2018-10	Pork
EC18PK-0040-2	PRJNA1177805	SRR31115747	SAMN44451987	101	B1	2018-10	Pork
EC18PK-0041-1	PRJNA1177805	SRR31115746	SAMN44451988	5474	B1	2018-10	Pork
EC18PK-0041-2	PRJNA1177805	SRR31115745	SAMN44451989	1204	D	2018-10	Pork
EC18PK-0058-1	PRJNA1177805	SRR31115744	SAMN44451990	971	B1	2018-10	Pork
EC18PK-0058-2	PRJNA1177805	SRR31115743	SAMN44451991	90	C	2018-10	Pork
EC18PK-0059-1	PRJNA1177805	SRR31115742	SAMN44451992	101	B1	2018-10	Pork
EC18PK-0059-2	PRJNA1177805	SRR31115741	SAMN44451993	409	A	2018-10	Pork
EC18PK-0064-1	PRJNA1177805	SRR31115740	SAMN44451994	4247	E	2018-10	Pork
EC18PK-0064-2	PRJNA1177805	SRR31115738	SAMN44451995	4247	E	2018-10	Pork
EC18PK-0065-1	PRJNA1177805	SRR31115737	SAMN44451996	6745	E	2018-10	Pork
EC18PK-0065-2	PRJNA1177805	SRR31115736	SAMN44451997	10	A	2018-10	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0066-1	PRJNA1177805	SRR31115735	SAMN44451998	101	B1	2018-10	Pork
EC18PK-0067-1	PRJNA1177805	SRR31115734	SAMN44451999	131	B2	2018-10	Pork
EC18PK-0067-2	PRJNA1177805	SRR31115733	SAMN44452000	131	B2	2018-10	Pork
EC18PK-0069-1	PRJNA1177805	SRR31115732	SAMN44452001	131	B2	2018-10	Pork
EC18PK-0069-2	PRJNA1177805	SRR31115731	SAMN44452002	131	B2	2018-10	Pork
EC18PK-0070-1	PRJNA1177805	SRR31115730	SAMN44452003	25 Novel	B1	2018-10	Pork
EC18PK-0070-2	PRJNA1177805	SRR31115729	SAMN44452004	25 Novel	B1	2018-10	Pork
EC18PK-0071-1	PRJNA1177805	SRR31115727	SAMN44452005	26 Novel	B1	2018-10	Pork
EC18PK-0071-2	PRJNA1177805	SRR31115726	SAMN44452006	26 Novel	B1	2018-10	Pork
EC18PK-0073-1	PRJNA1177805	SRR31115725	SAMN44452007	27 Novel	B1	2018-11	Pork
EC18PK-0073-2	PRJNA1177805	SRR31115724	SAMN44452008	27 Novel	B1	2018-11	Pork
EC18PK-0074-1	PRJNA1177805	SRR31115947	SAMN44452009	401	A	2018-11	Pork
EC18PK-0074-2	PRJNA1177805	SRR31115946	SAMN44452010	401	A	2018-11	Pork
EC18PK-0075-1	PRJNA1177805	SRR31115945	SAMN44452011	361	A	2018-11	Pork
EC18PK-0075-2	PRJNA1177805	SRR31115944	SAMN44452012	10	A	2018-11	Pork
EC18PK-0076-1	PRJNA1177805	SRR31115943	SAMN44452013	13	B1	2018-11	Pork
EC18PK-0076-2	PRJNA1177805	SRR31115942	SAMN44452014	13	B1	2018-11	Pork
EC18PK-0077-1	PRJNA1177805	SRR31115940	SAMN44452015	10	A	2018-11	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0077-2	PRJNA1177805	SRR31115939	SAMN44452016	10	A	2018-11	Pork
EC18PK-0078-1	PRJNA1177805	SRR31115938	SAMN44452017	10	A	2018-11	Pork
EC18PK-0078-2	PRJNA1177805	SRR31115937	SAMN44452018	28 Novel	A	2018-11	Pork
EC18PK-0079-1	PRJNA1177805	SRR31115936	SAMN44452019	1244	A	2018-11	Pork
EC18PK-0079-2	PRJNA1177805	SRR31115935	SAMN44452020	1244	A	2018-11	Pork
EC18PK-0080-2	PRJNA1177805	SRR31115934	SAMN44452021	6150	A	2018-11	Pork
EC18PK-0081-1	PRJNA1177805	SRR31115933	SAMN44452022	348	B1	2018-11	Pork
EC18PK-0081-2	PRJNA1177805	SRR31115932	SAMN44452023	348	B1	2018-11	Pork
EC18PK-0082-1	PRJNA1177805	SRR31115931	SAMN44452024	10	A	2018-11	Pork
EC18PK-0082-2	PRJNA1177805	SRR31115929	SAMN44452025	10	A	2018-11	Pork
EC18PK-0084-1	PRJNA1177805	SRR31115928	SAMN44452026	345	B1	2018-11	Pork
EC18PK-0084-2	PRJNA1177805	SRR31115927	SAMN44452027	847	B1	2018-11	Pork
EC18PK-0085-1	PRJNA1177805	SRR31115926	SAMN44452028	58	B1	2018-11	Pork
EC18PK-0085-2	PRJNA1177805	SRR31115925	SAMN44452029	58	B1	2018-11	Pork
EC18PK-0086-1	PRJNA1177805	SRR31115924	SAMN44452030	101	B1	2018-11	Pork
EC18PK-0087-1	PRJNA1177805	SRR31115923	SAMN44452031	75	B1	2018-11	Pork
EC18PK-0087-2	PRJNA1177805	SRR31115922	SAMN44452032	75	B1	2018-11	Pork
EC18PK-0088-1	PRJNA1177805	SRR31115921	SAMN44452033	10	A	2018-11	Pork
EC18PK-0088-2	PRJNA1177805	SRR31115920	SAMN44452034	24 Novel	A	2018-11	Pork
EC18PK-0089-1	PRJNA1177805	SRR31115918	SAMN44452035	165	A	2018-11	Pork
EC18PK-0089-2	PRJNA1177805	SRR31115917	SAMN44452036	453	B1	2018-11	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0090-1	PRJNA1177805	SRR31115916	SAMN44452037	101	B1	2018-11	Pork
EC18PK-0090-2	PRJNA1177805	SRR31115418	SAMN44452038	101	B1	2018-11	Pork
EC18PK-0091-1	PRJNA1177805	SRR31115417	SAMN44452039	101	B1	2018-11	Pork
EC18PK-0091-2	PRJNA1177805	SRR31115416	SAMN44452040	101	B1	2018-11	Pork
EC18PK-0098-1	PRJNA1177805	SRR31115415	SAMN44452041	2628	B1	2018-12	Pork
EC18PK-0098-2	PRJNA1177805	SRR31115414	SAMN44452042	4580	B1	2018-12	Pork
EC18PK-0099-1	PRJNA1177805	SRR31115413	SAMN44452043	216	A	2018-12	Pork
EC18PK-0099-2	PRJNA1177805	SRR31115412	SAMN44452044	4994	cladeI	2018-12	Pork
EC18PK-0107-1	PRJNA1177805	SRR31115410	SAMN44452045	216	A	2018-12	Pork
EC18PK-0119-2	PRJNA1177805	SRR31115409	SAMN44452046	29 Novel	A	2019-01	Pork
EC18PK-0120-1	PRJNA1177805	SRR31115408	SAMN44452047	88	C	2019-01	Pork
EC18PK-0120-2	PRJNA1177805	SRR31115407	SAMN44452048	410	C	2019-01	Pork
EC18PK-0121-1	PRJNA1177805	SRR31115406	SAMN44452049	117	G	2019-01	Pork
EC18PK-0121-2	PRJNA1177805	SRR31115405	SAMN44452050	117	G	2019-01	Pork
EC18PK-0122-1	PRJNA1177805	SRR31115404	SAMN44452051	117	G	2019-01	Pork
EC18PK-0122-2	PRJNA1177805	SRR31115403	SAMN44452052	117	G	2019-01	Pork
EC18PK-0123-1	PRJNA1177805	SRR31115402	SAMN44452053	10	A	2019-01	Pork
EC18PK-0123-2	PRJNA1177805	SRR31115401	SAMN44452054	6778	A	2019-01	Pork
EC18PK-0124-1	PRJNA1177805	SRR31115399	SAMN44452055	30 Novel	B1	2019-01	Pork
EC18PK-0124-2	PRJNA1177805	SRR31115398	SAMN44452056	30 Novel	B1	2019-01	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0126-1	PRJNA1177805	SRR31115397	SAMN44452057	5474	B1	2019-01	Pork
EC18PK-0126-2	PRJNA1177805	SRR31115396	SAMN44452058	635	A	2019-01	Pork
EC18PK-0127-1	PRJNA1177805	SRR31115395	SAMN44452059	131	B2	2019-01	Pork
EC18PK-0127-2	PRJNA1177805	SRR31115394	SAMN44452060	101	B1	2019-01	Pork
EC18PK-0128-1	PRJNA1177805	SRR31115393	SAMN44452061	567	B2	2019-01	Pork
EC18PK-0128-2	PRJNA1177805	SRR31115392	SAMN44452062	10	A	2019-01	Pork
EC18PK-0130-1	PRJNA1177805	SRR31115391	SAMN44452063	448	B1	2019-01	Pork
EC18PK-0130-2	PRJNA1177805	SRR31115390	SAMN44452064	58	B1	2019-01	Pork
EC18PK-0131-1	PRJNA1177805	SRR31115387	SAMN44452065	31 Novel	B1	2019-01	Pork
EC18PK-0131-2	PRJNA1177805	SRR31115386	SAMN44452066	31 Novel	B1	2019-01	Pork
EC18PK-0132-1	PRJNA1177805	SRR31115385	SAMN44452067	31 Novel	B1	2019-01	Pork
EC18PK-0132-2	PRJNA1177805	SRR31115384	SAMN44452068	1308	B1	2019-01	Pork
EC18PK-0133-1	PRJNA1177805	SRR31115383	SAMN44452069	542	A	2019-01	Pork
EC18PK-0133-2	PRJNA1177805	SRR31115382	SAMN44452070	46	A	2019-01	Pork
EC18PK-0134-1	PRJNA1177805	SRR31115381	SAMN44452071	10	A	2019-01	Pork
EC18PK-0134-2	PRJNA1177805	SRR31115380	SAMN44452072	10	A	2019-01	Pork
EC18PK-0135-1	PRJNA1177805	SRR31115379	SAMN44452073	542	A	2019-01	Pork
EC18PK-0135-2	PRJNA1177805	SRR31115378	SAMN44452074	542	A	2019-01	Pork
EC18PK-0136-1	PRJNA1177805	SRR31115376	SAMN44452075	2705	A	2019-02	Pork
EC18PK-0136-2	PRJNA1177805	SRR31115375	SAMN44452076	2705	A	2019-02	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0137-1	PRJNA1177805	SRR31115374	SAMN44452077	58	B1	2019-02	Pork
EC18PK-0137-2	PRJNA1177805	SRR31115373	SAMN44452078	10	A	2019-02	Pork
EC18PK-0138-1	PRJNA1177805	SRR31115372	SAMN44452079	351	B1	2019-02	Pork
EC18PK-0140-1	PRJNA1177805	SRR31115371	SAMN44452080	446	B1	2019-02	Pork
EC18PK-0140-2	PRJNA1177805	SRR31115370	SAMN44452081	1431	B1	2019-02	Pork
EC18PK-0141-1	PRJNA1177805	SRR31115369	SAMN44452082	446	B1	2019-02	Pork
EC18PK-0141-2	PRJNA1177805	SRR31115368	SAMN44452083	446	B1	2019-02	Pork
EC18PK-0142-1	PRJNA1177805	SRR31115367	SAMN44452084	1434	A	2019-02	Pork
EC18PK-0142-2	PRJNA1177805	SRR31115365	SAMN44452085	1434	A	2019-02	Pork
EC18PK-0143-1	PRJNA1177805	SRR31115364	SAMN44452086	101	B1	2019-02	Pork
EC18PK-0143-2	PRJNA1177805	SRR31115363	SAMN44452087	101	B1	2019-02	Pork
EC18PK-0144-1	PRJNA1177805	SRR31115362	SAMN44452088	131	B2	2019-02	Pork
EC18PK-0148-1	PRJNA1177805	SRR31115361	SAMN44452089	58	B1	2019-02	Pork
EC18PK-0148-2	PRJNA1177805	SRR31115360	SAMN44452090	93	A	2019-02	Pork
EC18PK-0149-1	PRJNA1177805	SRR31115359	SAMN44452091	2705	A	2019-02	Pork
EC18PK-0149-2	PRJNA1177805	SRR31115358	SAMN44452092	847	B1	2019-02	Pork
EC18PK-0153-1	PRJNA1177805	SRR31115357	SAMN44452093	108	D	2019-02	Pork
EC18PK-0153-2	PRJNA1177805	SRR31115356	SAMN44452094	398	A	2019-02	Pork
EC18PK-0154-1	PRJNA1177805	SRR31115354	SAMN44452095	32 Novel	B1	2019-02	Pork
EC18PK-0154-2	PRJNA1177805	SRR31115353	SAMN44452096	32 Novel	B1	2019-02	Pork
EC18PK-0203-1	PRJNA1107692	SRR28893293	SAMN41200187	2967	A	2019-06	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0203-2	PRJNA1107692	SRR28893292	SAMN41200188	2967	A	2019-06	Pork
EC18PK-0203-3	PRJNA1107692	SRR28893291	SAMN41200189	2967	A	2019-06	Pork
EC18PK-0220-1	PRJNA1177805	SRR31115352	SAMN44452097	101	B1	2019-06	Pork
EC18PK-0220-2	PRJNA1177805	SRR31115351	SAMN44452098	101	B1	2019-06	Pork
EC18PK-0221-1	PRJNA1107692	SRR28893290	SAMN41200190	2035	A	2019-06	Pork
EC18PK-0221-2	PRJNA1107692	SRR28893289	SAMN41200191	33 Novel	A	2019-06	Pork
EC18PK-0221-3	PRJNA1107692	SRR28893288	SAMN41200192	5409	B1	2019-06	Pork
EC18PK-0221-4	PRJNA1107692	SRR28893287	SAMN41200193	898	A	2019-06	Pork
EC18PK-0224-1	PRJNA1177805	SRR31115350	SAMN44452099	540	A	2019-07	Pork
EC18PK-0224-2	PRJNA1177805	SRR31115349	SAMN44452100	847	B1	2019-07	Pork
EC18PK-0225-1	PRJNA1107692	SRR28893285	SAMN41200194	34 Novel	E	2019-07	Pork
EC18PK-0225-2	PRJNA1107692	SRR28893284	SAMN41200195	215	A	2019-07	Pork
EC18PK-0225-3	PRJNA1107692	SRR28893283	SAMN41200196	10	A	2019-07	Pork
EC18PK-0225-4	PRJNA1107692	SRR28893282	SAMN41200197	10	A	2019-07	Pork
EC18PK-0230-1	PRJNA1107692	SRR28893281	SAMN41200198	10	A	2019-07	Pork
EC18PK-0230-2	PRJNA1107692	SRR28893280	SAMN41200199	10	A	2019-07	Pork
EC18PK-0230-3	PRJNA1107692	SRR28893279	SAMN41200200	10	A	2019-07	Pork
EC18PK-0230-4	PRJNA1107692	SRR28893278	SAMN41200201	10	A	2019-07	Pork
EC18PK-0235-1	PRJNA1107692	SRR28893277	SAMN41200202	196	B1	2019-07	Pork
EC18PK-0235-2	PRJNA1107692	SRR28893276	SAMN41200203	847	B1	2019-07	Pork
EC18PK-0235-3	PRJNA1107692	SRR28893274	SAMN41200204	847	B1	2019-07	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0235-4	PRJNA1107692	SRR28893273	SAMN41200205	196	B1	2019-07	Pork
EC18PK-0240-1	PRJNA1107692	SRR28893272	SAMN41200206	165	A	2019-08	Pork
EC18PK-0240-2	PRJNA1107692	SRR28893271	SAMN41200207	165	A	2019-08	Pork
EC18PK-0240-3	PRJNA1107692	SRR28893270	SAMN41200208	165	A	2019-08	Pork
EC18PK-0240-4	PRJNA1107692	SRR28893269	SAMN41200209	165	A	2019-08	Pork
EC18PK-0242-1	PRJNA1107692	SRR28893268	SAMN41200210	10	A	2019-08	Pork
EC18PK-0242-2	PRJNA1107692	SRR28893267	SAMN41200211	10	A	2019-08	Pork
EC18PK-0242-3	PRJNA1107692	SRR28893266	SAMN41200212	48	A	2019-08	Pork
EC18PK-0242-4	PRJNA1107692	SRR28893233	SAMN41200213	10	A	2019-08	Pork
EC18PK-0243-1	PRJNA1177805	SRR31115348	SAMN44452101	35 Novel	A	2019-08	Pork
EC18PK-0243-2	PRJNA1177805	SRR31115347	SAMN44452102	35 Novel	A	2019-08	Pork
EC18PK-0250-1	PRJNA1177805	SRR31115346	SAMN44452103	10	A	2019-08	Pork
EC18PK-0250-2	PRJNA1177805	SRR31115345	SAMN44452104	641	B1	2019-08	Pork
EC18PK-0251-1	PRJNA1177805	SRR31115343	SAMN44452105	349	D	2019-08	Pork
EC18PK-0251-2	PRJNA1177805	SRR31115342	SAMN44452106	10	A	2019-08	Pork
EC18PK-0255-1	PRJNA1107692	SRR28893231	SAMN41200214	1434	A	2019-08	Pork
EC18PK-0258-1	PRJNA1177805	SRR31115341	SAMN44452107	472	A	2019-09	Pork
EC18PK-0259-1	PRJNA1177805	SRR31115340	SAMN44452108	641	B1	2019-09	Pork
EC18PK-0259-2	PRJNA1177805	SRR31115339	SAMN44452109	101	B1	2019-09	Pork
EC18PK-0264-2	PRJNA1177805	SRR31115338	SAMN44452110	101	B1	2019-09	Pork
EC18PK-0265-1	PRJNA1177805	SRR31115337	SAMN44452111	1308	B1	2019-09	Pork

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PK-0265-2	PRJNA1177805	SRR31115336	SAMN44452112	1308	B1	2019-09	Pork
EC18PK-0266-1	PRJNA1107692	SRR28893230	SAMN41200215	162	B1	2019-09	Pork
EC18PK-0266-2	PRJNA1107692	SRR28893229	SAMN41200216	2967	A	2019-09	Pork
EC18PK-0266-3	PRJNA1107692	SRR28893228	SAMN41200217	58	B1	2019-09	Pork
EC18PK-0267-1	PRJNA1177805	SRR31115335	SAMN44452113	2967	A	2019-09	Pork
EC18PK-0267-2	PRJNA1177805	SRR31115334	SAMN44452114	2967	A	2019-09	Pork
EC18PK-0272-1	PRJNA1107692	SRR28893227	SAMN41200218	1844	B1	2019-09	Pork
EC18PK-0272-2	PRJNA1107692	SRR28893226	SAMN41200219	1844	B1	2019-09	Pork
EC18PK-0272-3	PRJNA1107692	SRR28893225	SAMN41200220	1125	B1	2019-09	Pork
EC18PK-0275-1	PRJNA1107692	SRR28893224	SAMN41200221	410	C	2019-09	Pork
EC18PK-0275-2	PRJNA1107692	SRR28893223	SAMN41200222	410	C	2019-09	Pork
EC18PK-0280-1	PRJNA1107692	SRR28893222	SAMN41200223	1020	B1	2019-10	Pork
EC18PK-0280-2	PRJNA1107692	SRR28893219	SAMN41200224	131	B2	2019-10	Pork
EC18PK-0280-3	PRJNA1107692	SRR28893218	SAMN41200225	131	B2	2019-10	Pork
EC18PK-0285-2	PRJNA1177805	SRR31115332	SAMN44452115	24 Novel	A	2019-10	Pork
EC18PK-0290-2	PRJNA1177805	SRR31115331	SAMN44452116	361	A	2019-10	Pork
EC18PK-0292-1	PRJNA1107692	SRR28893217	SAMN41200226	164	B1	2019-10	Pork
EC18PK-0306-1	PRJNA1107692	SRR28893216	SAMN41200227	399	A	2019-11	Pork
EC18PK-0306-2	PRJNA1107692	SRR28893215	SAMN41200228	399	A	2019-11	Pork
EC18PR-0001-1	PRJNA1177805	SRR31115330	SAMN44452117	226	A	2018-05	Prawns
EC18PR-0001-2	PRJNA1177805	SRR31115329	SAMN44452118	1079	B1	2018-05	Prawns
EC18PR-0001-3	PRJNA1177805	SRR31115328	SAMN44452119	607	A	2018-05	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0001-4	PRJNA1177805	SRR31115327	SAMN44452120	607	A	2018-05	Prawns
EC18PR-0004-1	PRJNA1177805	SRR31115326	SAMN44452121	36 Novel	G	2018-05	Prawns
EC18PR-0004-2	PRJNA1177805	SRR31115325	SAMN44452122	36 Novel	G	2018-05	Prawns
EC18PR-0004-3	PRJNA1177805	SRR31115324	SAMN44452123	36 Novel	G	2018-05	Prawns
EC18PR-0004-4	PRJNA1177805	SRR31115323	SAMN44452124	37 Novel	A	2018-05	Prawns
EC18PR-0006-1	PRJNA1177805	SRR31115689	SAMN44452125	38 Novel	D	2018-07	Prawns
EC18PR-0006-2	PRJNA1177805	SRR31115688	SAMN44452126	38 Novel	D	2018-07	Prawns
EC18PR-0006-3	PRJNA1177805	SRR31115687	SAMN44452127	38 Novel	D	2018-07	Prawns
EC18PR-0006-4	PRJNA1177805	SRR31115686	SAMN44452128	38 Novel	D	2018-07	Prawns
EC18PR-0007-1	PRJNA1177805	SRR31115685	SAMN44452129	39 Novel	B1	2018-08	Prawns
EC18PR-0007-2	PRJNA1177805	SRR31115684	SAMN44452130	939	A	2018-08	Prawns
EC18PR-0007-3	PRJNA1177805	SRR31115683	SAMN44452131	1196	B1	2018-08	Prawns
EC18PR-0007-4	PRJNA1177805	SRR31115682	SAMN44452132	215	A	2018-08	Prawns
EC18PR-0008-1	PRJNA1177805	SRR31115681	SAMN44452133	48	A	2018-08	Prawns
EC18PR-0008-2	PRJNA1177805	SRR31115680	SAMN44452134	48	A	2018-08	Prawns
EC18PR-0008-3	PRJNA1177805	SRR31115678	SAMN44452135	216	A	2018-08	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0008-4	PRJNA1177805	SRR31115677	SAMN44452136	216	A	2018-08	Prawns
EC18PR-0009-1	PRJNA1177805	SRR31115676	SAMN44452137	10512	B1	2018-08	Prawns
EC18PR-0009-2	PRJNA1177805	SRR31115675	SAMN44452138	10512	B1	2018-08	Prawns
EC18PR-0009-3	PRJNA1177805	SRR31115674	SAMN44452139	10512	B1	2018-08	Prawns
EC18PR-0009-4	PRJNA1177805	SRR31115673	SAMN44452140	10512	B1	2018-08	Prawns
EC18PR-0011-2	PRJNA1177805	SRR31115672	SAMN44452141	156	B1	2018-09	Prawns
EC18PR-0014-1	PRJNA1177805	SRR31115671	SAMN44452142	216	A	2018-09	Prawns
EC18PR-0014-2	PRJNA1177805	SRR31115670	SAMN44452143	216	A	2018-09	Prawns
EC18PR-0014-3	PRJNA1177805	SRR31115669	SAMN44452144	40 Novel	A	2018-09	Prawns
EC18PR-0014-4	PRJNA1177805	SRR31115667	SAMN44452145	216	A	2018-09	Prawns
EC18PR-0015-1	PRJNA1177805	SRR31115666	SAMN44452146	5295	A	2018-09	Prawns
EC18PR-0015-2	PRJNA1177805	SRR31115665	SAMN44452147	5295	A	2018-09	Prawns
EC18PR-0015-3	PRJNA1177805	SRR31115664	SAMN44452148	5295	A	2018-09	Prawns
EC18PR-0015-4	PRJNA1177805	SRR31115663	SAMN44452149	5295	A	2018-09	Prawns
EC18PR-0016-1	PRJNA1177805	SRR31115662	SAMN44452150	5295	A	2018-09	Prawns
EC18PR-0016-2	PRJNA1177805	SRR31115661	SAMN44452151	5295	A	2018-09	Prawns
EC18PR-0016-3	PRJNA1177805	SRR31115660	SAMN44452152	5295	A	2018-09	Prawns
EC18PR-0016-4	PRJNA1177805	SRR31115659	SAMN44452153	5295	A	2018-09	Prawns
EC18PR-0018-1	PRJNA1177805	SRR31115658	SAMN44452154	41 Novel	B1	2018-09	Prawns
EC18PR-0018-2	PRJNA1177805	SRR31115656	SAMN44452155	41 Novel	B1	2018-09	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0018-3	PRJNA1177805	SRR31115655	SAMN44452156	41 Novel	B1	2018-09	Prawns
EC18PR-0018-4	PRJNA1177805	SRR31115654	SAMN44452157	41 Novel	B1	2018-09	Prawns
EC18PR-0024-1	PRJNA1177805	SRR31115653	SAMN44452158	181	A	2018-09	Prawns
EC18PR-0025-1	PRJNA1177805	SRR31115652	SAMN44452159	3580	B1	2018-09	Prawns
EC18PR-0025-2	PRJNA1177805	SRR31115651	SAMN44452160	2954	E	2018-09	Prawns
EC18PR-0025-3	PRJNA1177805	SRR31115650	SAMN44452161	3580	B1	2018-09	Prawns
EC18PR-0025-4	PRJNA1177805	SRR31115649	SAMN44452162	3580	B1	2018-09	Prawns
EC18PR-0026-1	PRJNA1177805	SRR31115648	SAMN44452163	216	A	2018-09	Prawns
EC18PR-0026-2	PRJNA1177805	SRR31115647	SAMN44452164	216	A	2018-09	Prawns
EC18PR-0026-3	PRJNA1177805	SRR31115644	SAMN44452165	216	A	2018-09	Prawns
EC18PR-0026-4	PRJNA1177805	SRR31115643	SAMN44452166	216	A	2018-09	Prawns
EC18PR-0027-1	PRJNA1177805	SRR31115642	SAMN44452167	216	A	2018-09	Prawns
EC18PR-0027-2	PRJNA1177805	SRR31115641	SAMN44452168	602	B1	2018-09	Prawns
EC18PR-0027-3	PRJNA1177805	SRR31115640	SAMN44452169	216	A	2018-09	Prawns
EC18PR-0027-4	PRJNA1177805	SRR31115639	SAMN44452170	602	B1	2018-09	Prawns
EC18PR-0033-1	PRJNA1177805	SRR31115638	SAMN44452171	216	A	2018-09	Prawns
EC18PR-0034-1	PRJNA1177805	SRR31115637	SAMN44452172	48	A	2018-09	Prawns
EC18PR-0034-2	PRJNA1177805	SRR31115636	SAMN44452173	224	B1	2018-09	Prawns
EC18PR-0034-3	PRJNA1177805	SRR31115635	SAMN44452174	224	B1	2018-09	Prawns
EC18PR-0034-4	PRJNA1177805	SRR31115633	SAMN44452175	42 Novel	B1	2018-09	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0035-1	PRJNA1177805	SRR31115632	SAMN44452176	86	B1	2018-09	Prawns
EC18PR-0035-2	PRJNA1177805	SRR31115631	SAMN44452177	6027	D	2018-09	Prawns
EC18PR-0035-3	PRJNA1177805	SRR31115630	SAMN44452178	6027	D	2018-09	Prawns
EC18PR-0035-4	PRJNA1177805	SRR31115629	SAMN44452179	6027	D	2018-09	Prawns
EC18PR-0037-1	PRJNA1177805	SRR31115628	SAMN44452180	58	A	2018-09	Prawns
EC18PR-0037-2	PRJNA1177805	SRR31115627	SAMN44452181	2160	B1	2018-09	Prawns
EC18PR-0037-3	PRJNA1177805	SRR31115723	SAMN44452182	43 Novel	A	2018-09	Prawns
EC18PR-0037-4	PRJNA1177805	SRR31115722	SAMN44452183	4684	B1	2018-09	Prawns
EC18PR-0044-1	PRJNA1177805	SRR31115721	SAMN44452184	607	A	2018-09	Prawns
EC18PR-0044-2	PRJNA1177805	SRR31115719	SAMN44452185	607	A	2018-09	Prawns
EC18PR-0048-1	PRJNA1177805	SRR31115718	SAMN44452186	2161	B1	2018-09	Prawns
EC18PR-0048-2	PRJNA1177805	SRR31115717	SAMN44452187	2161	B1	2018-09	Prawns
EC18PR-0051-1	PRJNA1177805	SRR31115716	SAMN44452188	48	A	2018-09	Prawns
EC18PR-0051-2	PRJNA1177805	SRR31115715	SAMN44452189	8576	A	2018-09	Prawns
EC18PR-0055-1	PRJNA1177805	SRR31115714	SAMN44452190	2169	B1	2018-09	Prawns
EC18PR-0055-2	PRJNA1177805	SRR31115713	SAMN44452191	2169	B1	2018-09	Prawns
EC18PR-0055-3	PRJNA1177805	SRR31115712	SAMN44452192	155	B1	2018-09	Prawns
EC18PR-0055-4	PRJNA1177805	SRR31115711	SAMN44452193	155	B1	2018-09	Prawns
EC18PR-0059-1	PRJNA1177805	SRR31115710	SAMN44452194	101	B1	2018-10	Prawns
EC18PR-0059-2	PRJNA1177805	SRR31115708	SAMN44452195	101	B1	2018-10	Prawns
EC18PR-0059-3	PRJNA1177805	SRR31115530	SAMN44452196	101	B1	2018-10	Prawns
EC18PR-0059-4	PRJNA1177805	SRR31115529	SAMN44452197	101	B1	2018-10	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0067-1	PRJNA1177805	SRR31115528	SAMN44452198	44 Novel	cladeIV	2018-10	Prawns
EC18PR-0067-2	PRJNA1177805	SRR31115527	SAMN44452199	45 Novel	A	2018-10	Prawns
EC18PR-0067-3	PRJNA1177805	SRR31115526	SAMN44452200	44 Novel	cladeIV	2018-10	Prawns
EC18PR-0067-4	PRJNA1177805	SRR31115525	SAMN44452201	44 Novel	cladeIV	2018-10	Prawns
EC18PR-0071-1	PRJNA1177805	SRR31115524	SAMN44452202	683	B1	2018-10	Prawns
EC18PR-0071-2	PRJNA1177805	SRR31115523	SAMN44452203	683	B1	2018-10	Prawns
EC18PR-0075-1	PRJNA1177805	SRR31115522	SAMN44452204	216	A	2018-10	Prawns
EC18PR-0075-2	PRJNA1177805	SRR31115520	SAMN44452205	216	A	2018-10	Prawns
EC18PR-0076-1	PRJNA1177805	SRR31115519	SAMN44452206	942	B1	2018-10	Prawns
EC18PR-0076-2	PRJNA1177805	SRR31115518	SAMN44452207	58	B1	2018-10	Prawns
EC18PR-0078-1	PRJNA1177805	SRR31115517	SAMN44452208	86	B1	2018-10	Prawns
EC18PR-0078-2	PRJNA1177805	SRR31115516	SAMN44452209	86	B1	2018-10	Prawns
EC18PR-0078-3	PRJNA1177805	SRR31115515	SAMN44452210	86	B1	2018-10	Prawns
EC18PR-0078-4	PRJNA1177805	SRR31115626	SAMN44452211	5359	A	2018-10	Prawns
EC18PR-0081-1	PRJNA1177805	SRR31115625	SAMN44452212	3232	B1	2018-10	Prawns
EC18PR-0081-2	PRJNA1177805	SRR31115624	SAMN44452213	3232	B1	2018-10	Prawns
EC18PR-0081-3	PRJNA1177805	SRR31115623	SAMN44452214	3232	B1	2018-10	Prawns
EC18PR-0081-4	PRJNA1177805	SRR31115621	SAMN44452215	3232	B1	2018-10	Prawns
EC18PR-0082-1	PRJNA1177805	SRR31115620	SAMN44452216	3858	B1	2018-10	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0082-2	PRJNA1177805	SRR31115619	SAMN44452217	3858	B1	2018-10	Prawns
EC18PR-0082-3	PRJNA1177805	SRR31115618	SAMN44452218	3858	B1	2018-10	Prawns
EC18PR-0082-4	PRJNA1177805	SRR31115617	SAMN44452219	3858	B1	2018-10	Prawns
EC18PR-0086-1	PRJNA1177805	SRR31115616	SAMN44452220	5523	A	2018-10	Prawns
EC18PR-0086-2	PRJNA1177805	SRR31115615	SAMN44452221	5523	A	2018-10	Prawns
EC18PR-0086-3	PRJNA1177805	SRR31115614	SAMN44452222	5523	A	2018-10	Prawns
EC18PR-0086-4	PRJNA1177805	SRR31115613	SAMN44452223	5523	A	2018-10	Prawns
EC18PR-0090-1	PRJNA1177805	SRR31115612	SAMN44452224	46 Novel	A	2018-10	Prawns
EC18PR-0090-2	PRJNA1177805	SRR31115610	SAMN44452225	46 Novel	A	2018-10	Prawns
EC18PR-0094-1	PRJNA1177805	SRR31115609	SAMN44452226	3489	A	2018-10	Prawns
EC18PR-0094-2	PRJNA1177805	SRR31115608	SAMN44452227	3489	A	2018-10	Prawns
EC18PR-0094-3	PRJNA1177805	SRR31115607	SAMN44452228	3489	A	2018-10	Prawns
EC18PR-0094-4	PRJNA1177805	SRR31115606	SAMN44452229	3489	A	2018-10	Prawns
EC18PR-0097-1	PRJNA1177805	SRR31115605	SAMN44452230	10	A	2018-10	Prawns
EC18PR-0097-2	PRJNA1177805	SRR31115604	SAMN44452231	10	A	2018-10	Prawns
EC18PR-0097-3	PRJNA1177805	SRR31115603	SAMN44452232	10	A	2018-10	Prawns
EC18PR-0097-4	PRJNA1177805	SRR31115602	SAMN44452233	10	A	2018-10	Prawns
EC18PR-0107-1	PRJNA1177805	SRR31115601	SAMN44452234	156	B1	2018-10	Prawns
EC18PR-0107-2	PRJNA1177805	SRR31115599	SAMN44452235	224	B1	2018-10	Prawns
EC18PR-0109-1	PRJNA1177805	SRR31115598	SAMN44452236	5523	A	2018-11	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0109-2	PRJNA1177805	SRR31115597	SAMN44452237	5523	A	2018-11	Prawns
EC18PR-0110-1	PRJNA1177805	SRR31115596	SAMN44452238	86	B1	2018-11	Prawns
EC18PR-0110-2	PRJNA1177805	SRR31115595	SAMN44452239	86	B1	2018-11	Prawns
EC18PR-0114-1	PRJNA1177805	SRR31115915	SAMN44452240	5523	A	2018-11	Prawns
EC18PR-0114-2	PRJNA1177805	SRR31115914	SAMN44452241	5523	A	2018-11	Prawns
EC18PR-0116-1	PRJNA1177805	SRR31115913	SAMN44452242	101	B1	2018-11	Prawns
EC18PR-0116-2	PRJNA1177805	SRR31115912	SAMN44452243	101	B1	2018-11	Prawns
EC18PR-0118-1	PRJNA1177805	SRR31115911	SAMN44452244	1079	B1	2018-11	Prawns
EC18PR-0118-2	PRJNA1177805	SRR31115909	SAMN44452245	1079	B1	2018-11	Prawns
EC18PR-0119-1	PRJNA1177805	SRR31115908	SAMN44452246	47 Novel	B1	2018-11	Prawns
EC18PR-0119-2	PRJNA1177805	SRR31115907	SAMN44452247	47 Novel	B1	2018-11	Prawns
EC18PR-0123-1	PRJNA1177805	SRR31115906	SAMN44452248	399	A	2018-11	Prawns
EC18PR-0123-2	PRJNA1177805	SRR31115905	SAMN44452249	399	A	2018-11	Prawns
EC18PR-0125-1	PRJNA1177805	SRR31115904	SAMN44452250	357	B2	2018-11	Prawns
EC18PR-0125-2	PRJNA1177805	SRR31115903	SAMN44452251	357	B2	2018-11	Prawns
EC18PR-0126-1	PRJNA1177805	SRR31115902	SAMN44452252	4162	B1	2018-11	Prawns
EC18PR-0126-2	PRJNA1177805	SRR31115901	SAMN44452253	1823	A	2018-11	Prawns
EC18PR-0131-1	PRJNA1177805	SRR31115900	SAMN44452254	155	B1	2018-11	Prawns
EC18PR-0131-2	PRJNA1177805	SRR31115898	SAMN44452255	1421	A	2018-11	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0132-1	PRJNA1177805	SRR31115897	SAMN44452256	48 Novel	B1	2018-11	Prawns
EC18PR-0132-2	PRJNA1177805	SRR31115896	SAMN44452257	48 Novel	B1	2018-11	Prawns
EC18PR-0133-1	PRJNA1177805	SRR31115895	SAMN44452258	155	B1	2018-11	Prawns
EC18PR-0137-1	PRJNA1177805	SRR31115894	SAMN44452259	4682	B1	2018-11	Prawns
EC18PR-0137-2	PRJNA1177805	SRR31115893	SAMN44452260	4682	B1	2018-11	Prawns
EC18PR-0143-1	PRJNA1177805	SRR31115892	SAMN44452261	49 Novel	E	2018-11	Prawns
EC18PR-0143-2	PRJNA1177805	SRR31115891	SAMN44452262	1049	B1	2018-11	Prawns
EC18PR-0150-1	PRJNA1177805	SRR31115890	SAMN44452263	1148	B1	2018-11	Prawns
EC18PR-0150-2	PRJNA1177805	SRR31115889	SAMN44452264	1148	B1	2018-11	Prawns
EC18PR-0151-1	PRJNA1177805	SRR31115886	SAMN44452265	155	B1	2018-11	Prawns
EC18PR-0151-2	PRJNA1177805	SRR31115885	SAMN44452266	448	B1	2018-11	Prawns
EC18PR-0154-1	PRJNA1177805	SRR31115884	SAMN44452267	1434	A	2018-12	Prawns
EC18PR-0154-2	PRJNA1177805	SRR31115322	SAMN44452268	1434	A	2018-12	Prawns
EC18PR-0157-1	PRJNA1177805	SRR31115321	SAMN44452269	1434	A	2018-12	Prawns
EC18PR-0157-2	PRJNA1177805	SRR31115320	SAMN44452270	1434	A	2018-12	Prawns
EC18PR-0160-1	PRJNA1177805	SRR31115319	SAMN44452271	50 Novel	A	2018-12	Prawns
EC18PR-0160-2	PRJNA1177805	SRR31115318	SAMN44452272	50 Novel	A	2018-12	Prawns
EC18PR-0163-1	PRJNA1177805	SRR31115317	SAMN44452273	10	A	2018-12	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0163-2	PRJNA1177805	SRR31115316	SAMN44452274	10	A	2018-12	Prawns
EC18PR-0164-1	PRJNA1177805	SRR31115314	SAMN44452275	215	A	2018-12	Prawns
EC18PR-0164-2	PRJNA1177805	SRR31115313	SAMN44452276	215	A	2018-12	Prawns
EC18PR-0166-1	PRJNA1177805	SRR31115312	SAMN44452277	10	A	2018-12	Prawns
EC18PR-0166-2	PRJNA1177805	SRR31115311	SAMN44452278	10	A	2018-12	Prawns
EC18PR-0170-1	PRJNA1177805	SRR31115310	SAMN44452279	399	A	2018-12	Prawns
EC18PR-0170-2	PRJNA1177805	SRR31115309	SAMN44452280	399	A	2018-12	Prawns
EC18PR-0172-1	PRJNA1177805	SRR31115308	SAMN44452281	409	A	2018-12	Prawns
EC18PR-0172-2	PRJNA1177805	SRR31115307	SAMN44452282	409	A	2018-12	Prawns
EC18PR-0173-1	PRJNA1177805	SRR31115306	SAMN44452283	155	B1	2018-12	Prawns
EC18PR-0173-2	PRJNA1177805	SRR31115305	SAMN44452284	155	B1	2018-12	Prawns
EC18PR-0174-1	PRJNA1177805	SRR31115303	SAMN44452285	155	B1	2018-12	Prawns
EC18PR-0174-2	PRJNA1177805	SRR31115302	SAMN44452286	1246	B1	2018-12	Prawns
EC18PR-0175-1	PRJNA1177805	SRR31115301	SAMN44452287	8262	D	2018-12	Prawns
EC18PR-0175-2	PRJNA1177805	SRR31115300	SAMN44452288	3232	B1	2018-12	Prawns
EC18PR-0177-1	PRJNA1177805	SRR31115299	SAMN44452289	2521	B1	2018-12	Prawns
EC18PR-0177-2	PRJNA1177805	SRR31115298	SAMN44452290	51 Novel	A	2018-12	Prawns
EC18PR-0178-1	PRJNA1177805	SRR31115297	SAMN44452291	4221	E	2018-12	Prawns
EC18PR-0178-2	PRJNA1177805	SRR31115296	SAMN44452292	4221	E	2018-12	Prawns
EC18PR-0179-1	PRJNA1177805	SRR31115295	SAMN44452293	48	A	2018-12	Prawns
EC18PR-0179-2	PRJNA1177805	SRR31115294	SAMN44452294	48	A	2018-12	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0183-1	PRJNA1177805	SRR31115292	SAMN44452295	46	A	2019-01	Prawns
EC18PR-0183-2	PRJNA1177805	SRR31115291	SAMN44452296	52 Novel	A	2019-01	Prawns
EC18PR-0184-1	PRJNA1177805	SRR31115514	SAMN44452297	3863	B1	2019-01	Prawns
EC18PR-0184-2	PRJNA1177805	SRR31115513	SAMN44452298	3863	B1	2019-01	Prawns
EC18PR-0189-1	PRJNA1177805	SRR31115512	SAMN44452299	197	A	2019-01	Prawns
EC18PR-0189-2	PRJNA1177805	SRR31115511	SAMN44452300	197	A	2019-01	Prawns
EC18PR-0189-3	PRJNA1177805	SRR31115510	SAMN44452301	197	A	2019-01	Prawns
EC18PR-0189-4	PRJNA1177805	SRR31115509	SAMN44452302	197	A	2019-01	Prawns
EC18PR-0190-1	PRJNA1177805	SRR31115508	SAMN44452303	939	A	2019-01	Prawns
EC18PR-0190-2	PRJNA1177805	SRR31115507	SAMN44452304	939	A	2019-01	Prawns
EC18PR-0190-3	PRJNA1177805	SRR31115505	SAMN44452305	939	A	2019-01	Prawns
EC18PR-0190-4	PRJNA1177805	SRR31115504	SAMN44452306	2825	B1	2019-01	Prawns
EC18PR-0191-1	PRJNA1177805	SRR31115503	SAMN44452307	10	A	2019-01	Prawns
EC18PR-0191-2	PRJNA1177805	SRR31115502	SAMN44452308	10	A	2019-01	Prawns
EC18PR-0191-3	PRJNA1177805	SRR31115501	SAMN44452309	10	A	2019-01	Prawns
EC18PR-0191-4	PRJNA1177805	SRR31115500	SAMN44452310	10	A	2019-01	Prawns
EC18PR-0197-1	PRJNA1177805	SRR31115499	SAMN44452311	53 Novel	D	2019-02	Prawns
EC18PR-0197-2	PRJNA1177805	SRR31115498	SAMN44452312	10	A	2019-02	Prawns
EC18PR-0197-3	PRJNA1177805	SRR31115497	SAMN44452313	4377	D	2019-02	Prawns
EC18PR-0197-4	PRJNA1177805	SRR31115496	SAMN44452314	4377	D	2019-02	Prawns

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18PR-0251-1	PRJNA1107692	SRR28893214	SAMN41200229	2531	A	2019-08	Prawns
EC18PR-0251-2	PRJNA1107692	SRR28893213	SAMN41200230	2531	A	2019-08	Prawns
EC18PR-0251-3	PRJNA1107692	SRR28893212	SAMN41200231	2531	A	2019-08	Prawns
EC18PR-0251-4	PRJNA1107692	SRR28893211	SAMN41200232	2531	A	2019-08	Prawns
EC18PR-0256-1	PRJNA1107692	SRR28893210	SAMN41200233	6241	B1	2019-09	Prawns
EC18PR-0256-2	PRJNA1107692	SRR28893208	SAMN41200234	6241	B1	2019-09	Prawns
EC18PR-0256-3	PRJNA1107692	SRR28893207	SAMN41200235	6241	B1	2019-09	Prawns
EC18PR-0262-1	PRJNA1107692	SRR28893206	SAMN41200236	409	A	2019-09	Prawns
EC18SM-0006-1	PRJNA1177805	SRR31115494	SAMN44452315	409	A	2018-07	Salmon
EC18SM-0006-2	PRJNA1177805	SRR31115493	SAMN44452316	409	A	2018-07	Salmon
EC18SM-0006-3	PRJNA1177805	SRR31115492	SAMN44452317	409	A	2018-07	Salmon
EC18SM-0006-4	PRJNA1177805	SRR31115491	SAMN44452318	409	A	2018-07	Salmon
EC18SM-0007-1	PRJNA1177805	SRR31115490	SAMN44452319	6163	G	2018-08	Salmon
EC18SM-0007-2	PRJNA1177805	SRR31115489	SAMN44452320	6163	F	2018-08	Salmon
EC18SM-0007-3	PRJNA1177805	SRR31115488	SAMN44452321	6163	F	2018-08	Salmon
EC18SM-0007-4	PRJNA1177805	SRR31115487	SAMN44452322	6163	F	2018-08	Salmon
EC18SM-0013-1	PRJNA1177805	SRR31115486	SAMN44452323	5474	B1	2018-09	Salmon
EC18SM-0013-2	PRJNA1177805	SRR31115485	SAMN44452324	5474	B1	2018-09	Salmon
EC18SM-0013-3	PRJNA1177805	SRR31115483	SAMN44452325	5474	B1	2018-09	Salmon
EC18SM-0013-4	PRJNA1177805	SRR31115594	SAMN44452326	5474	B1	2018-09	Salmon
EC18SM-0015-1	PRJNA1177805	SRR31115593	SAMN44452327	5474	B1	2018-09	Salmon

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18SM-0015-2	PRJNA1177805	SRR31115592	SAMN44452328	5474	B1	2018-09	Salmon
EC18SM-0015-3	PRJNA1177805	SRR31115591	SAMN44452329	448	B1	2018-09	Salmon
EC18SM-0015-4	PRJNA1177805	SRR31115590	SAMN44452330	5474	B1	2018-09	Salmon
EC18SM-0021-1	PRJNA1177805	SRR31115589	SAMN44452331	541	A	2018-09	Salmon
EC18SM-0021-2	PRJNA1177805	SRR31115588	SAMN44452332	541	A	2018-09	Salmon
EC18SM-0021-3	PRJNA1177805	SRR31115587	SAMN44452333	541	A	2018-09	Salmon
EC18SM-0021-4	PRJNA1177805	SRR31115586	SAMN44452334	541	A	2018-09	Salmon
EC18SM-0028-1	PRJNA1177805	SRR31115584	SAMN44452335	720	D	2018-10	Salmon
EC18SM-0028-2	PRJNA1177805	SRR31115583	SAMN44452336	720	D	2018-10	Salmon
EC18SM-0029-1	PRJNA1177805	SRR31115582	SAMN44452337	86	B1	2018-10	Salmon
EC18SM-0029-2	PRJNA1177805	SRR31115581	SAMN44452338	5474	B1	2018-10	Salmon
EC18SM-0042-1	PRJNA1177805	SRR31115580	SAMN44452339	399	A	2018-11	Salmon
EC18SM-0049-1	PRJNA1177805	SRR31115579	SAMN44452340	5474	B1	2018-11	Salmon
EC18SM-0049-2	PRJNA1177805	SRR31115578	SAMN44452341	5474	B1	2018-11	Salmon
EC18SM-0054-1	PRJNA1177805	SRR31115577	SAMN44452342	1115	A	2018-12	Salmon
EC18SM-0054-2	PRJNA1177805	SRR31115576	SAMN44452343	1115	A	2018-12	Salmon
EC18SM-0054-3	PRJNA1177805	SRR31115575	SAMN44452344	1115	A	2018-12	Salmon
EC18SM-0054-4	PRJNA1177805	SRR31115573	SAMN44452345	1115	A	2018-12	Salmon
EC18SM-0059-1	PRJNA1177805	SRR31115572	SAMN44452346	155	B1	2019-01	Salmon
EC18SM-0059-2	PRJNA1177805	SRR31115571	SAMN44452347	155	B1	2019-01	Salmon
EC18SM-0060-1	PRJNA1177805	SRR31115570	SAMN44452348	5474	B1	2019-01	Salmon

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18SM-0060-2	PRJNA1177805	SRR31115569	SAMN44452349	5474	B1	2019-01	Salmon
EC18SM-0067-1	PRJNA1177805	SRR31115568	SAMN44452350	5474	B1	2019-01	Salmon
EC18SM-0067-2	PRJNA1177805	SRR31115567	SAMN44452351	5474	B1	2019-01	Salmon
EC18SM-0067-3	PRJNA1177805	SRR31115566	SAMN44452352	5474	B1	2019-01	Salmon
EC18SM-0067-4	PRJNA1177805	SRR31115565	SAMN44452353	5474	B1	2019-01	Salmon
EC18SM-0069-1	PRJNA1177805	SRR31115564	SAMN44452354	54 Novel	A	2019-01	Salmon
EC18SM-0069-2	PRJNA1177805	SRR31115883	SAMN44452355	8763	A	2019-01	Salmon
EC18SM-0069-3	PRJNA1177805	SRR31115882	SAMN44452356	8763	A	2019-01	Salmon
EC18SM-0069-4	PRJNA1177805	SRR31115881	SAMN44452357	8763	A	2019-01	Salmon
EC18SM-0075-1-1	PRJNA1177805	SRR31115880	SAMN44452358	1434	A	2019-02	Salmon
EC18SM-0075-2-1	PRJNA1177805	SRR31115879	SAMN44452359	1434	A	2019-02	Salmon
EC18SM-0075-3	PRJNA1177805	SRR31115878	SAMN44452360	1434	A	2019-02	Salmon
EC18SM-0075-4	PRJNA1177805	SRR31115877	SAMN44452361	1434	A	2019-02	Salmon
EC18SM-0078-1-1	PRJNA1177805	SRR31115876	SAMN44452362	8763	A	2019-02	Salmon
EC18SM-0078-2	PRJNA1177805	SRR31115875	SAMN44452363	8763	A	2019-02	Salmon
EC18SM-0078-3	PRJNA1177805	SRR31115874	SAMN44452364	8763	A	2019-02	Salmon
EC18SM-0078-4	PRJNA1177805	SRR31115871	SAMN44452365	8763	A	2019-02	Salmon
EC18SM-0082-1-1	PRJNA1177805	SRR31115870	SAMN44452366	1434	A	2019-04	Salmon

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18SM-0082-2-1	PRJNA1177805	SRR31115869	SAMN44452367	1434	A	2019-04	Salmon
EC18SM-0082-3	PRJNA1177805	SRR31115868	SAMN44452368	1434	A	2019-04	Salmon
EC18SM-0082-4	PRJNA1177805	SRR31115867	SAMN44452369	1434	A	2019-04	Salmon
EC18SM-0084-1	PRJNA1177805	SRR31115866	SAMN44452370	8763	A	2019-04	Salmon
EC18SM-0088-1	PRJNA1177805	SRR31115865	SAMN44452371	1146	B1	2019-04	Salmon
EC18SM-0088-2	PRJNA1177805	SRR31115864	SAMN44452372	1146	B1	2019-04	Salmon
EC18SM-0105-1	PRJNA1107692	SRR28893205	SAMN41200237	720	D	2019-06	Salmon
EC18SM-0105-2-R	PRJNA1177805	SRR31115863	SAMN44452373	720	D	2019-06	Salmon
EC18SM-0105-3-R	PRJNA1177805	SRR31115862	SAMN44452374	720	D	2019-06	Salmon
EC18SM-0105-4-R	PRJNA1177805	SRR31115860	SAMN44452375	720	D	2019-06	Salmon
EC18SM-0112-1	PRJNA1177805	SRR31115859	SAMN44452376	2967	A	2019-06	Salmon
EC18SM-0112-2	PRJNA1177805	SRR31115858	SAMN44452377	2967	A	2019-06	Salmon
EC18SM-0112-3	PRJNA1177805	SRR31115857	SAMN44452378	2967	A	2019-06	Salmon
EC18SM-0112-4	PRJNA1177805	SRR31115856	SAMN44452379	2967	A	2019-06	Salmon
EC18SM-0116-1	PRJNA1177805	SRR31115855	SAMN44452380	24 Novel	A	2019-07	Salmon
EC18SM-0116-2	PRJNA1177805	SRR31115854	SAMN44452381	24 Novel	A	2019-07	Salmon
EC18SM-0116-3	PRJNA1177805	SRR31115853	SAMN44452382	24 Novel	A	2019-07	Salmon

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18SM-0116-4	PRJNA1177805	SRR31115852	SAMN44452383	24 Novel	A	2019-07	Salmon
EC18SM-0119-1	PRJNA1177805	SRR31115562	SAMN44452384	24 Novel	A	2019-07	Salmon
EC18SM-0119-2	PRJNA1177805	SRR31115560	SAMN44452385	24 Novel	A	2019-07	Salmon
EC18SM-0119-3	PRJNA1177805	SRR31115559	SAMN44452386	24 Novel	A	2019-07	Salmon
EC18SM-0119-4	PRJNA1177805	SRR31115558	SAMN44452387	24 Novel	A	2019-07	Salmon
EC18SM-0120-1	PRJNA1107692	SRR28893416	SAMN41200241	536	A	2019-07	Salmon
EC18SM-0120-2	PRJNA1107692	SRR28893415	SAMN41200242	536	A	2019-07	Salmon
EC18SM-0120-3	PRJNA1107692	SRR28893414	SAMN41200243	536	A	2019-07	Salmon
EC18SM-0120-4	PRJNA1107692	SRR28893412	SAMN41200244	536	A	2019-07	Salmon
EC18SM-0121-1	PRJNA1177805	SRR31115557	SAMN44452388	536	A	2019-07	Salmon
EC18SM-0121-2	PRJNA1177805	SRR31115556	SAMN44452389	536	A	2019-07	Salmon
EC18SM-0121-3	PRJNA1177805	SRR31115555	SAMN44452390	536	A	2019-07	Salmon
EC18SM-0121-4	PRJNA1177805	SRR31115554	SAMN44452391	536	A	2019-07	Salmon
EC18SM-0124-1	PRJNA1107692	SRR28893411	SAMN41200245	480	A	2019-08	Salmon
EC18SM-0124-2	PRJNA1107692	SRR28893410	SAMN41200246	480	A	2019-08	Salmon
EC18SM-0124-3	PRJNA1107692	SRR28893409	SAMN41200247	480	A	2019-08	Salmon
EC18SM-0125-1	PRJNA1177805	SRR31115553	SAMN44452392	5474	B1	2019-08	Salmon
EC18SM-0125-2	PRJNA1177805	SRR31115552	SAMN44452393	5474	B1	2019-08	Salmon

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18SM-0125-3	PRJNA1177805	SRR31115551	SAMN44452394	5474	B1	2019-08	Salmon
EC18SM-0125-4	PRJNA1177805	SRR31115549	SAMN44452395	216	A	2019-08	Salmon
EC18SM-0127-1	PRJNA1177805	SRR31115548	SAMN44452396	216	A	2019-08	Salmon
EC18SM-0127-2	PRJNA1177805	SRR31115547	SAMN44452397	216	A	2019-08	Salmon
EC18SM-0127-3	PRJNA1177805	SRR31115546	SAMN44452398	24 Novel	A	2019-08	Salmon
EC18SM-0127-4	PRJNA1177805	SRR31115545	SAMN44452399	216	A	2019-08	Salmon
EC18SM-0130-1	PRJNA1177805	SRR31115544	SAMN44452400	216	A	2019-09	Salmon
EC18SM-0130-2	PRJNA1177805	SRR31115543	SAMN44452401	216	A	2019-09	Salmon
EC18SM-0130-3	PRJNA1177805	SRR31115542	SAMN44452402	399	A	2019-09	Salmon
EC18SM-0130-4	PRJNA1177805	SRR31115541	SAMN44452403	216	A	2019-09	Salmon
EC18SM-0135-1-R	PRJNA1177805	SRR31115540	SAMN44452404	3604	B1	2019-09	Salmon
EC18SM-0135-2-R	PRJNA1177805	SRR31115538	SAMN44452405	3604	B1	2019-09	Salmon
EC18SM-0135-3-R	PRJNA1177805	SRR31115537	SAMN44452406	3604	B1	2019-09	Salmon
EC18SM-0137-1	PRJNA1177805	SRR31115536	SAMN44452407	1429	A	2019-09	Salmon
EC18SM-0137-2	PRJNA1177805	SRR31115535	SAMN44452408	1429	A	2019-09	Salmon
EC18SM-0149-1	PRJNA1107692	SRR28893405	SAMN41200251	355	B2	2019-10	Salmon
EC18SM-0149-2	PRJNA1107692	SRR28893404	SAMN41200252	355	B2	2019-10	Salmon
EC18SM-0149-3-R	PRJNA1177805	SRR31115534	SAMN44452409	355	B2	2019-10	Salmon

Appendix 2: Accession information, multilocus sequence types, phylogroups, date collected and food commodity metadata of the 1,067 complete *Escherichia coli* genomes generated within this study

Genome	Bioproject accession	Accession	Biosample Accession	ST	Phylogroup	Date Collected	Commodity
EC18SM-0149-4	PRJNA1107692	SRR28893401	SAMN41200254	355	B2	2019-10	Salmon
EC18SM-0152-1	PRJNA1177805	SRR31115533	SAMN44452410	3604	B1	2019-11	Salmon
EC18SM-0152-2	PRJNA1177805	SRR31115532	SAMN44452411	3604	B1	2019-11	Salmon
EC18SM-0152-3	PRJNA1177805	SRR31115531	SAMN44452412	3604	B1	2019-11	Salmon
EC18SM-0152-4	PRJNA1177805	SRR31115290	SAMN44452413	3604	B1	2019-11	Salmon

Appendix 3: Allelic profiles of the novel sequence types (ST) identified within this thesis.

Novel ST	*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*	Thesis chapter
Novel 1	adk(6)	fumC(11)	gyrB(4)	icd(~876)	mdh(7)	purA(8)	recA(6)	Chapter 4
Novel 2	adk(83)	fumC(186)	gyrB(136)	icd(12)	mdh(1)	purA(2)	recA(2)	Chapter 4
Novel 3	adk(~6)	fumC(6)	gyrB(5)	icd(10)	mdh(9)	purA(8)	recA(6)	Chapter 4
Novel 4	adk(6)	fumC(19)	gyrB(3)	icd(18)	mdh(9)	purA(13,13)	recA(156)	Chapter 4
Novel 5	adk(~20)	fumC(45)	gyrB(41)	icd(43)	mdh(5)	purA(32)	recA(2)	Chapter 4
Novel 6	adk(6)	fumC(4)	gyrB(33)	icd(132)	mdh(20)	purA(~186)	recA(7)	Chapter 4
Novel 7	adk(6)	fumC(23)	gyrB(608)	icd(16)	mdh(27)	purA(35)	recA(6)	Chapter 4
Novel 8	adk(1065)	fumC(11)	gyrB(4)	icd(8)	mdh(8)	purA(8)	recA(2)	Chapter 4
Novel 9	adk(6)	fumC(4)	gyrB(5)	icd(18)	mdh(~11)	purA(8)	recA(14)	Chapter 4
Novel 10	adk(10)	fumC(11)	gyrB(4)	icd(8)	mdh(8)	purA(382)	recA(369)	Chapter 4
Novel 11	adk(1)	fumC(251)	gyrB(207)	icd(25)	mdh(180)	purA(5)	recA(~19)	Chapter 4
Novel 12	adk(1063?)	fumC(4)	gyrB(3)	icd(16)	mdh(11)	purA(8)	recA(6)	Chapter 4

Novel 13	adk(6)	fumC(65)	gyrB(32)	icd(~26)	mdh(11)	purA(8)	recA(2)	Chapter 4
Novel 14	adk(6)	fumC(65)	gyrB(33)	icd(16)	mdh(7)	purA(8)	recA(6)	Chapter 4
Novel 15	adk(6)	fumC(4)	gyrB(4)	icd(16)	mdh(24)	purA(~8)	recA(14)	Chapter 4
Novel 16	adk(6)	fumC(~11)	gyrB(4)	icd(8)	mdh(8)	purA(8)	recA(2)	Chapter 4
Novel 17	adk(10)	fumC(11)	gyrB(4)	icd(8)	mdh(616)	purA(8)	recA(2)	Chapter 4
Novel 18	adk(6)	fumC(~93)	gyrB(26)	icd(82)	mdh(1)	purA(2)	recA(2)	Chapter 4
Novel 19	adk(6)	fumC(19)	gyrB(22)	icd(26)	mdh(11)	purA(8)	recA(2)	Chapter 4
Novel 20	adk(6)	fumC(95)	gyrB(15)	icd(18)	mdh(9)	purA(8)	recA(6)	Chapter 4
Novel 21	adk(224)	fumC(4)	gyrB(54)	icd(247)	mdh(61)	purA(1)	recA(7)	Chapter 4
Novel 22	adk(876?)	fumC(45)	gyrB(41)	icd(43)	mdh(5)	purA(32)	recA(2)	Chapter 4
Novel 23	adk(10)	fumC(~11)	gyrB(57)	icd(8)	mdh(7)	purA(18)	recA(6)	Chapter 4
Novel 24	adk(8)	fumC(107)	gyrB(4)	icd(8)	mdh(8)	purA(8)	recA(2)	Chapter 4
Novel 25	adk(6)	fumC(~29)	gyrB(32)	icd(16)	mdh(9)	purA(8)	recA(2)	Chapter 4
Novel 26	adk(6)	fumC(6)	gyrB(5)	icd(85)	mdh(9)	purA(8)	recA(7)	Chapter 4
Novel 27	adk(6)	fumC(6)	gyrB(~33)	icd(1)	mdh(24)	purA(7)	recA(7)	Chapter 4
Novel 28	adk(~6)	fumC(4)	gyrB(1)	icd(95)	mdh(69)	purA(8)	recA(20)	Chapter 4
Novel 29	adk(10)	fumC(~594)	gyrB(5)	icd(8)	mdh(7)	purA(8)	recA(6)	Chapter 4
Novel 30	adk(6)	fumC(6)	gyrB(5)	icd(10)	mdh(20)	purA(~23)	recA(6)	Chapter 4
Novel 31	adk(6)	fumC(6)	gyrB(33)	icd(1084)	mdh(24)	purA(8)	recA(7)	Chapter 4
Novel 32	adk(6)	fumC(~19)	gyrB(3)	icd(18)	mdh(9)	purA(13)	recA(156)	Chapter 4
Novel 33	adk(10)	fumC(~7)	gyrB(4)	icd(8)	mdh(12)	purA(8)	recA(2)	Chapter 4
Novel 34	adk(87)	fumC(~958)	gyrB(~186)	icd(~1123)	mdh(~42)	purA(162)	recA(~440)	Chapter 4
Novel 35	adk(10)	fumC(7?)	gyrB(4)	icd(8)	mdh(12)	purA(8)	recA(2)	Chapter 4
Novel 36	adk(20)	fumC(~1466)	gyrB(41)	icd(43)	mdh(~5)	purA(548)	recA(46)	Chapter 4
Novel 6	adk(100)	fumC(23)	gyrB(7)	icd(45)	mdh(7)	purA(7)	recA(7)	Chapter 5

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0001-1	aph_3____Ib, aph_6__Id, tet_A
EC18CH-0001-2	
EC18CH-0001-3	blaTEM
EC18CH-0001-4	
EC18CH-0002-1	
EC18CH-0002-2	
EC18CH-0002-3	aph_3____Ib, aph_6__Id, tet_B
EC18CH-0002-4	aph_3____Ib, aph_6__Id, tet_B
EC18CH-0003-1	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, lnu_F, qnrS1, sul2, tet_A
EC18CH-0003-2	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, lnu_F, qnrS1, sul2, tet_A
EC18CH-0003-3	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, lnu_F, qnrS1, sul2, tet_A
EC18CH-0003-4	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, lnu_F, qnrS1, sul2, tet_A
EC18CH-0004-1	
EC18CH-0004-2	aadA1, blaTEM, dfrA1, sul1, tet_A
EC18CH-0004-3	aadA1, blaTEM, dfrA1, sul1, tet_A
EC18CH-0005-1	tet_A
EC18CH-0005-2	
EC18CH-0005-3	
EC18CH-0005-4	tet_A
EC18CH-0006-1	
EC18CH-0006-2	
EC18CH-0006-3	23S r.754G>A, aph_3____Ib, aph_6__Id, tet_B
EC18CH-0006-4	blaTEM, dfrA1, sul2, tet_A
EC18CH-0007-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, gyrA_S83L, sul2, tet_A
EC18CH-0007-2	blaTEM_220, gyrA_S83L, tet_A
EC18CH-0007-3	blaTEM_220, gyrA_S83L, tet_A
EC18CH-0007-4	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, gyrA_S83L, sul2, tet_A
EC18CH-0008-1	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, dfrA14, qnrB19, sul2, tet_A, tet_B
EC18CH-0008-2	aadA2, aadA5, blaTEM, dfrA17, lnu_F, sul2, tet_B
EC18CH-0008-3	blaTEM_220, gyrA_S83L, tet_A
EC18CH-0008-4	
EC18CH-0009-1	
EC18CH-0009-2	aadA1, aph_3___Ia, blaTEM, dfrA1, sul1, tet_A
EC18CH-0009-3	aac_3__IV, aadA2, aph_4__Ia, blaTEM, lnu_F, sul2, tet_A
EC18CH-0010-1	blaTEM, sul1, sul2
EC18CH-0010-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0010-3	parC_S57T
EC18CH-0010-4	aadA1, lnu_G, sul2, tet_B
EC18CH-0011-1	
EC18CH-0011-2	aadA1, blaTEM, dfrA1, sul1, tet_A
EC18CH-0011-3	aadA1, blaTEM, dfrA1, sul1, tet_A
EC18CH-0011-4	tet_A, tet_M
EC18CH-0012-1	gyrA_S83L
EC18CH-0012-2	aadA1, blaTEM_30, dfrA1, floR, gyrA_D87N, gyrA_S83L, parC_S80I, sul2, tet_B
EC18CH-0012-4	
EC18CH-0013-1	aac_3__IId, aadA5, aph_3____Ib, aph_6__Id, blaTEM, dfrA17, gyrA_D87N, gyrA_S83L, mph_A, parC_S80I, sul1, sul2, tet_B
EC18CH-0013-2	aac_3__IId, aadA5, aph_3____Ib, aph_6__Id, blaTEM, dfrA17, gyrA_D87N, gyrA_S83L, mph_A, parC_S80I, sul1, sul2, tet_B
EC18CH-0013-4	aadA1, blaTEM, dfrA1, gyrA_D87N, gyrA_S83L, parC_S80I, sul1, tet_A
EC18CH-0014-1	aadA1, aph_3____Ib, aph_6__Id, dfrA1, mph_B, sul1, sul2, tet_A
EC18CH-0014-2	blaTEM, dfrA1, sul2
EC18CH-0014-3	
EC18CH-0014-4	
EC18CH-0015-1	
EC18CH-0015-2	
EC18CH-0015-3	
EC18CH-0016-1	blaTEM
EC18CH-0017-1	aadA1, aadA2, blaTEM, cmlA1, dfrA12, gyrA_D87N, gyrA_S83L, parC_S80I, sul3, tet_A
EC18CH-0018-1	aadA1, blaTEM, sul1
EC18CH-0019-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, sul2, tet_A
EC18CH-0022-1	
EC18CH-0023-1	aadA1, blaTEM, dfrA1, gyrA_S83L, sul2
EC18CH-0025-4	aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0026-1	tet_A
EC18CH-0026-3	
EC18CH-0028-2	aadA22, aph_3____Ib, aph_3___Ia, aph_6__Id, blaTEM, lnu_F, sul2, tet_B
EC18CH-0029-2	blaTEM
EC18CH-0029-3	aph_3____Ib, aph_6__Id
EC18CH-0029-4	dfrA5
EC18CH-0030-2	aadA1, sul1

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0030-3	aph_3____Ib, aph_6__Id, tet_A, tet_B
EC18CH-0030-4	aadA22, aph_6__Id, lnu_F, tet_B
EC18CH-0033-2	fosA7, parE_I355T
EC18CH-0040-1	
EC18CH-0040-2	
EC18CH-0041-1	fosA7
EC18CH-0041-2	
EC18CH-0058-1	aph_3____Ib, aph_6__Id, blaTEM, sul2, tet_A
EC18CH-0058-2	aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA17, sul2, tet_B
EC18CH-0059-1	sul2
EC18CH-0059-2	blaTEM
EC18CH-0066-1	
EC18CH-0066-2	
EC18CH-0067-1	blaTEM, dfrA1, dfrA5, sul2
EC18CH-0067-2	
EC18CH-0068-1	
EC18CH-0068-2	
EC18CH-0076-1	blaTEM, tet_B
EC18CH-0076-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, lnu_G, sul2, tet_A
EC18CH-0077-1	
EC18CH-0077-2	sul2
EC18CH-0080-1	aadA1, sul1
EC18CH-0080-2	
EC18CH-0081-1	aadA1, blaTEM, dfrA1, sul2
EC18CH-0081-2	aadA5, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, dfrA17, lnu_F, sul2, tet_B
EC18CH-0083-1	
EC18CH-0083-2	aadA1, blaTEM, dfrA1, tet_A
EC18CH-0090-1	
EC18CH-0090-2	
EC18CH-0091-1	
EC18CH-0091-2	
EC18CH-0092-1	blaTEM, gyrA_D87G, gyrA_S83L, parC_S80I, tet_A
EC18CH-0092-2	aph_3____Ib, aph_3___Ia, aph_6__Id, blaTEM, catA1, gyrA_D87N, gyrA_S83L, parC_A56T, parC_S80I, sul2, tet_B
EC18CH-0093-1	blaTEM
EC18CH-0093-2	
EC18CH-0094-1	sul2
EC18CH-0094-2	aph_3____Ib, aph_6__Id, blaTEM, gyrA_S83L, parC_S80I, sul2

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0099-1	
EC18CH-0099-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0100-1	aph_3____Ib, aph_6__Id
EC18CH-0100-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0103-1	tet_B
EC18CH-0103-2	dfrA14, qnrS1, sul2, tet_A
EC18CH-0106-1	aadA17, blaTEM, dfrA1, lnu_F, sul2
EC18CH-0106-2	aadA1, blaTEM, dfrA1, lnu_G
EC18CH-0107-1	aadA1, blaTEM, dfrA1, sul2
EC18CH-0107-2	aadA1, blaTEM, dfrA1, sul2
EC18CH-0120-1	
EC18CH-0120-2	blaTEM, dfrA1, sul2, tet_A
EC18CH-0121-1	
EC18CH-0121-2	blaTEM, dfrA1, sul2
EC18CH-0122-1	aph_3____Ib, aph_6__Id, blaTEM
EC18CH-0122-2	aph_3____Ib, aph_6__Id, tet_A
EC18CH-0123-2	qnrB19
EC18CH-0124-2	aadA22, gyrA_S83L, lnu_F, parE_I355T
EC18CH-0126-1	aac_3__IId, aadA5, aph_3____Ib, aph_6__Id, blaTEM, dfrA17, fosA7, gyrA_D87N, gyrA_S83L, mph_A, parC_S80I, sul1, sul2, tet_A
EC18CH-0127-2	aph_3____Ib, blaTEM, floR, gyrA_D87N, gyrA_S83L, parC_S80I, sul2, tet_A
EC18CH-0128-1	aadA24, blaTEM, dfrA1, lnu_G, sul2, tet_A
EC18CH-0128-2	aac_3__VIa, aadA1, sul1, tet_C
EC18CH-0130-1	
EC18CH-0130-2	
EC18CH-0132-1	blaTEM
EC18CH-0132-2	blaTEM
EC18CH-0133-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, gyrA_D87N, gyrA_S83L, parC_E84G, parC_S80I, sul2
EC18CH-0135-2	
EC18CH-0137-2	aadA1, aph_3____Ib, blaTEM, dfrA1, gyrA_D87N, gyrA_S83L, lnu_F, parC_S80I, sul1, sul2, tet_A
EC18CH-0138-1	
EC18CH-0138-2	
EC18CH-0139-1	blaTEM, sul3
EC18CH-0139-2	
EC18CH-0140-1	blaTEM

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0140-2	blaTEM
EC18CH-0141-2	aadA5, dfrA17, gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458T
EC18CH-0144-1	
EC18CH-0144-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, sul2
EC18CH-0148-1	aadA1, blaOXA_10, catB3, dfrB4, sul1
EC18CH-0148-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0149-1	aadA17, blaTEM, cmlA1, lnu_F, sul3, tet_A
EC18CH-0149-2	blaTEM, dfrA1, lnu_G, sul2, tet_A
EC18CH-0154-1	aph_3___Ia, dfrA5
EC18CH-0154-2	aph_3___Ia, dfrA5
EC18CH-0155-1	aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0155-2	
EC18CH-0176-1	
EC18CH-0176-2	
EC18CH-0192-1	blaTEM, dfrA1, qnrS1, sul2, tet_A
EC18CH-0192-2	blaTEM, dfrA1, sul2
EC18CH-0192-3	aadA1, dfrA1, sul1, tet_A
EC18CH-0192-4	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, gyrA_S83L, parC_S80I, sul2
EC18CH-0195-1	
EC18CH-0195-2	aadA1, aph_3___Ia, blaTEM, dfrA1, sul1, tet_A
EC18CH-0195-3	aph_3____Ib, aph_6__Id, blaTEM, gyrA_S83L, sul2, tet_A
EC18CH-0195-4	qnrB19
EC18CH-0201-1	
EC18CH-0201-2	gyrA_S83L
EC18CH-0201-3	
EC18CH-0201-4	
EC18CH-0202-1	aph_3____Ib, aph_6__Id, sul2, tet_A
EC18CH-0202-2	blaTEM, dfrA1, sul2
EC18CH-0202-3	aph_3____Ib, aph_6__Id, sul2, tet_A
EC18CH-0212-1	aadA2, blaTEM, lnu_F, tet_A
EC18CH-0212-2	aadA2, blaTEM, lnu_F, tet_A
EC18CH-0212-3	tet_A
EC18CH-0212-4	tet_A
EC18CH-0215-1	aac_3__IId, aadA2, blaTEM, dfrA17, gyrA_D87N, gyrA_S83L, lnu_F, parC_E84G, parC_S80I, parE_I355T, tet_B
EC18CH-0215-2	
EC18CH-0215-3	aac_3__IId, aadA2, blaTEM, dfrA17, gyrA_D87N, gyrA_S83L, lnu_F, parC_E84G, parC_S80I, parE_I355T, tet_B

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0215-4	
EC18CH-0216-1	fosA7, parE_I355T
EC18CH-0216-2	
EC18CH-0216-3	tet_A
EC18CH-0216-4	tet_A
EC18CH-0222-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, sul2
EC18CH-0222-2	qnrB19
EC18CH-0223-1	aadA1, blaTEM, dfrA1, gyrA_S83L, parC_S80I, sul2
EC18CH-0223-2	23S r.754G>A, aadA1, blaTEM, dfrA1, gyrA_S83L, parC_S80I, sul2
EC18CH-0223-3	blaTEM
EC18CH-0223-4	23S r.754G>A, aadA1, blaTEM, dfrA1, gyrA_S83L, parC_S80I, sul2
EC18CH-0224-1	
EC18CH-0224-2	
EC18CH-0224-3	
EC18CH-0224-4	
EC18CH-0225-1	
EC18CH-0225-2	
EC18CH-0232-1	
EC18CH-0232-2	sul2
EC18CH-0232-3	tet_A
EC18CH-0232-4	
EC18CH-0234-1	
EC18CH-0234-2	
EC18CH-0235-1	aph_3____Ib, aph_6__Id, tet_B
EC18CH-0235-2	aph_3____Ib, aph_6__Id, tet_B
EC18CH-0235-3	aph_3____Ib, aph_6__Id, tet_B
EC18CH-0235-4	blaTEM
EC18CH-0242-1	blaTEM, lnu_G
EC18CH-0242-2	aadA1, blaOXA_10, catB3, dfrB4, sul1
EC18CH-0243-1	blaTEM
EC18CH-0243-2	aph_6__Id, blaTEM, dfrA14, sul2
EC18CH-0243-3	
EC18CH-0243-4	dfrA5
EC18CH-0245-1	blaTEM, qnrB19, tet_A
EC18CH-0245-2	aadA5, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA17, floR, gyrA_D87N, gyrA_S83L, parC_S80I, sul2, tet_B
EC18CH-0245-3	blaTEM, gyrA_D87N, gyrA_S83L, parC_E84K
EC18CH-0245-4	blaTEM, dfrA36, gyrA_D87N, sul2, tet_A
EC18CH-0248-1	aadA1, dfrA1, sul1, tet_A

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0248-2	aadA1, aph_3____Ib, aph_6__Id, blaCTX_M_27, dfrA1, floR, sul1, sul2, tet_A, tet_B
EC18CH-0249-1	aph_3____Ib, aph_6__Id, blaTEM, sul2, tet_A
EC18CH-0249-2	blaTEM, dfrA36, qnrS1, sul2, tet_B
EC18CH-0251-1	blaTEM, dfrA1, sul2, tet_A
EC18CH-0251-2	blaTEM, dfrA1, sul2, tet_A
EC18CH-0251-3	aadA1, blaTEM, tet_A
EC18CH-0251-4	tet_A
EC18CH-0252-2	aph_3____Ib, aph_3___Ia, aph_6__Id, sul2, tet_B
EC18CH-0253-2	blaTEM, dfrA1, sul2, tet_B
EC18CH-0254-1	blaTEM, tet_B
EC18CH-0254-2	
EC18CH-0255-1	aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0255-2	aadA1, blaTEM_220, cmlA1, dfrA15, mcr_1, qnrS1, sul3, tet_A
EC18CH-0256-1	
EC18CH-0256-2	
EC18CH-0257-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, sul2
EC18CH-0257-2	blaTEM, dfrA7
EC18CH-0258-1	blaTEM, tet_A
EC18CH-0258-2	blaTEM, tet_A
EC18CH-0259-2	
EC18CH-0260-1	
EC18CH-0260-2	blaTEM, dfrA1, sul2, tet_A
EC18CH-0261-1	
EC18CH-0261-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, sul2, tet_A
EC18CH-0262-1	
EC18CH-0262-2	
EC18CH-0263-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, sul2
EC18CH-0263-2	aph_3____Ib, aph_6__Id, tet_B
EC18CH-0264-1	
EC18CH-0264-2	gyrA_D87N, gyrA_S83L, parC_S80I, tet_A
EC18CH-0264-3	aadA1, blaTEM, dfrA1, gyrA_D87N, gyrA_S83L, parC_S80I, sul1, tet_A
EC18CH-0264-4	blaTEM, gyrA_S83L, qnrB19
EC18CH-0265-1	aac_3__IId, aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, sul1, sul2, tet_A
EC18CH-0265-2	aadA1, dfrA1, sul2, tet_B
EC18CH-0266-1	aph_3____Ib, aph_3___Ia, aph_6__Id, blaTEM, sul2, tet_B
EC18CH-0266-2	blaTEM, qnrS1
EC18CH-0266-3	sul2

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0266-4	aph_3____Ib, aph_3___Ia, aph_6__Id, blaTEM, sul2, tet_B
EC18CH-0267-1	aadA1, dfrA1, sul1, tet_A
EC18CH-0267-2	aadA1, dfrA1, sul1, tet_A
EC18CH-0269-1	aadA1, blaTEM, dfrA1, sul2, tet_B
EC18CH-0269-2	aadA1, blaTEM_30, dfrA1, floR, sul1, sul2, tet_A
EC18CH-0270-1	
EC18CH-0270-2	blaTEM, qnrS1
EC18CH-0270-3	
EC18CH-0270-4	
EC18CH-0271-1	aadA5, aph_3____Ib, aph_6__Id, blaTEM, dfrA17, sul2, tet_B
EC18CH-0271-2	blaTEM, lnu_G
EC18CH-0273-1	
EC18CH-0273-2	
EC18CH-0273-3	
EC18CH-0273-4	
EC18CH-0274-1	
EC18CH-0274-2	
EC18CH-0275-1	aadA5, aph_3____Ib, aph_6__Id, blaTEM, dfrA17, sul2, tet_B
EC18CH-0275-2	blaTEM_40
EC18CH-0276-1	aadA1, aadA2, blaTEM, cmlA1, dfrA8, gyrA_S83L, parC_S80R, sul2, sul3
EC18CH-0276-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM, lnu_F, sul2, tet_B
EC18CH-0276-3	aadA1, blaTEM, dfrA1, gyrA_S83L, lnu_G, sul3, tet_A
EC18CH-0276-4	aadA1, aph_3____Ib, aph_3___Ia, aph_6__Id, blaTEM, dfrA14, erm_42, floR, gyrA_S83L, sul2, sul3, tet_A
EC18CH-0279-1	
EC18CH-0279-2	
EC18CH-0279-3	
EC18CH-0279-4	
EC18CH-0280-1	
EC18CH-0280-2	
EC18CH-0280-3	blaTEM
EC18CH-0281-1	aadA1, blaTEM, dfrA1, sul2
EC18CH-0282-1	
EC18CH-0282-2	
EC18CH-0282-3	
EC18CH-0282-4	blaTEM
EC18CH-0283-2	aadA5, aph_3____Ib, aph_6__Id, blaTEM, dfrA17, sul2, tet_A

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18CH-0288-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA1, gyrA_D87N, gyrA_S83L, mcr_1, parC_A56T, parC_S80I, sul2
EC18CH-0288-2	aadA1, blaTEM, dfrA1, qnrB19, sul2, tet_A
EC18CH-0288-3	aph_3____Ib, aph_6__Id, blaTEM, gyrA_D87N, gyrA_S83L, mcr_1, parC_A56T, parC_S80I, sul2
EC18CH-0288-4	blaTEM, fosA7, gyrA_D87N, gyrA_S83L, parC_S80I
EC18CH-0293-1	
EC18CH-0293-2	
EC18CH-0293-3	
EC18CH-0293-4	
EC18CH-0302-1	
EC18CH-0302-2	
EC18CH-0302-3	
EC18CH-0302-4	
EC18CH-0306-1	aadA1, blaTEM, dfrA1, sul2
EC18CH-0306-2	aadA1, blaTEM, dfrA1, sul2
EC18CH-0306-3	aadA1, blaTEM, dfrA1, sul2
EC18CH-0306-4	aph_3____Ib, aph_6__Id, blaTEM, parC_A56T
EC18CH-0310-1	
EC18CH-0310-2	blaTEM, tet_A
EC18CH-0310-3	
EC18CH-0311-1	aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, dfrA14, gyrA_D87N, gyrA_S83L, parC_S80I, sul1, sul2, tet_A
EC18CH-0311-2	blaTEM, gyrA_S83L
EC18CH-0311-3	aph_3____Ib, aph_6__Id, blaTEM, floR, gyrA_S83L, sul2, tet_B
EC18CH-0311-4	aadA1, blaTEM, cmlA1, dfrA1, dfrA15, qnrS1, sul2, sul3, tet_A
EC18LG-0001-1	
EC18LG-0001-2	
EC18LG-0001-3	
EC18LG-0001-4	
EC18LG-0003-1	
EC18LG-0003-2	
EC18LG-0003-3	
EC18LG-0003-4	
EC18LG-0005-1	
EC18LG-0005-2	
EC18LG-0010-1	aadA1, aadA2, blaCARB_2, cmlA1, dfrA16, gyrA_S83L, parC_S80I, sul3, tet_A
EC18LG-0010-2	aadA1, aadA2, blaCARB_2, cmlA1, dfrA16, gyrA_S83L, parC_S80I, sul3, tet_A
EC18LG-0010-3	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18LG-0010-4	aadA1, aadA2, blaCARB_2, cmlA1, dfrA16, gyrA_S83L, parC_S80I, sul3, tet_A
EC18LG-0011-1	
EC18LG-0011-2	
EC18LG-0011-3	
EC18LG-0011-4	
EC18LG-0013-1	
EC18LG-0013-2	
EC18LG-0013-3	
EC18LG-0013-4	
EC18LG-0014-1	
EC18LG-0014-2	
EC18LG-0014-3	
EC18LG-0014-4	
EC18LG-0016-1	
EC18LG-0016-2	parC_S57T
EC18LG-0016-3	
EC18LG-0016-4	
EC18LG-0017-1	blaCMY_2, parE_I355T
EC18LG-0017-2	blaCMY_2, parE_I355T
EC18LG-0017-3	blaCMY_2, parE_I355T
EC18LG-0017-4	blaCMY_2
EC18LG-0018-1	
EC18LG-0018-2	
EC18LG-0018-3	
EC18LG-0018-4	
EC18LG-0019-1	blaCMY_2
EC18LG-0020-1	
EC18LG-0020-2	
EC18LG-0020-3	
EC18LG-0020-4	
EC18LG-0021-1	
EC18LG-0021-2	
EC18LG-0021-3	
EC18LG-0021-4	
EC18LG-0022-1	
EC18LG-0024-1	
EC18LG-0024-2	
EC18LG-0024-3	
EC18LG-0024-4	
EC18LG-0026-1	
EC18LG-0026-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18LG-0026-3	
EC18LG-0026-4	
EC18LG-0028-1	
EC18LG-0028-2	
EC18LG-0028-3	
EC18LG-0028-4	
EC18LG-0029-1	
EC18LG-0029-2	
EC18LG-0029-3	
EC18LG-0029-4	
EC18LG-0030-1	
EC18LG-0030-2	
EC18LG-0030-3	blaTEM
EC18LG-0030-4	blaTEM
EC18LG-0032-1	
EC18LG-0032-2	
EC18LG-0032-3	
EC18LG-0032-4	
EC18LG-0037-1	
EC18LG-0037-2	
EC18LG-0037-3	
EC18LG-0037-4	
EC18LG-0039-2	
EC18LG-0039-4	
EC18LG-0041-1	
EC18LG-0041-2	
EC18LG-0041-3	
EC18LG-0041-4	
EC18LG-0043-1	
EC18LG-0043-2	
EC18LG-0043-3	
EC18LG-0043-4	
EC18LG-0044-1	
EC18LG-0044-2	
EC18LG-0044-3	
EC18LG-0044-4	
EC18LG-0045-1	
EC18LG-0045-2	
EC18LG-0045-3	
EC18LG-0045-4	
EC18LG-0071-1	
EC18LG-0071-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18LG-0096-1	parE_I529L
EC18LG-0097-1	gyrA_S83L, parC_S80I, tet_M
EC18LG-0097-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, floR, parC_S57T, sul2, tet_A
EC18LG-0100-1	
EC18LG-0102-1	
EC18LG-0102-2	
EC18LG-0203-1	tet_A
EC18LG-0207-1	
EC18LG-0207-2	
EC18LG-0207-3	
EC18LG-0207-4	
EC18LG-0208-1	
EC18LG-0208-2	
EC18LG-0208-3	
EC18LG-0211-1	
EC18LG-0211-2	
EC18LG-0211-3	
EC18LG-0211-4	aph_3____Ib
EC18LG-0216-1	
EC18LG-0216-2	
EC18LG-0216-3	
EC18LG-0218-1	
EC18LG-0218-2	
EC18LG-0218-3	
EC18LG-0218-4	
EC18LG-0222-1	
EC18LG-0222-2	parC_S57T, parE_I355T
EC18LG-0222-3	
EC18LG-0233-1	
EC18LG-0233-2	
EC18LG-0233-3	
EC18LG-0239-1	
EC18LG-0239-2	
EC18LG-0239-3	
EC18LG-0239-4	
EC18LG-0247-1	
EC18LG-0250-1	tet_A
EC18LG-0250-2	tet_A
EC18LG-0254-1	
EC18LG-0254-2	
EC18LG-0254-3	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18LG-0265-1	
EC18LG-0265-2	
EC18LG-0265-3	
EC18LG-0265-4	
EC18LG-0271-1	
EC18LG-0271-2	
EC18LG-0274-1	
EC18LG-0274-2	
EC18LG-0274-3	
EC18LG-0274-4	
EC18LG-0277-1	
EC18LG-0283-1	
EC18LG-0283-2	
EC18LG-0283-3	
EC18LG-0283-4	
EC18LG-0287-1	tet_A
EC18LG-0287-2	tet_A
EC18LG-0287-3	tet_A
EC18LG-0297-1	fosA7
EC18LG-0297-2	aph_3____Ib, aph_6__Id, qnrB19, tet_A
EC18LG-0297-3	
EC18LG-0298-1	
EC18LG-0298-2	
EC18LG-0302-1	
EC18LG-0302-2	
EC18PK-0002-1	
EC18PK-0002-2	aadA1, aadA2, aph_3____Ib, aph_6__Id, dfrA12, sul2
EC18PK-0002-3	
EC18PK-0002-4	
EC18PK-0003-1	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0003-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0003-3	
EC18PK-0003-4	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0007-1	aadA1, aadA2, blaTEM, cmlA1, dfrA1, sul1, sul3, tet_A
EC18PK-0007-2	aadA1, aadA2, blaTEM, cmlA1, dfrA1, sul1, sul3, tet_A
EC18PK-0007-3	gyrA_S83L
EC18PK-0007-4	gyrA_S83L
EC18PK-0008-1	aac_3__IId, aadA1, aph_3____Ib, aph_6__Id, blaTEM, dfrA1, gyrA_D87N, gyrA_S83L, parC_S80I, sul1, sul2, tet_A

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0008-2	aadA1, blaTEM, dfrA1, dfrA14, floR, qnrS1, sul2, tet_A
EC18PK-0008-3	blaTEM_220, gyrA_S83L, tet_A
EC18PK-0008-4	aadA1, blaTEM, dfrA1, dfrA14, floR, qnrS1, sul2, tet_A
EC18PK-0010-1	
EC18PK-0010-2	
EC18PK-0010-3	
EC18PK-0010-4	
EC18PK-0011-1	
EC18PK-0011-2	
EC18PK-0011-3	
EC18PK-0011-4	
EC18PK-0012-1	
EC18PK-0012-2	
EC18PK-0012-3	
EC18PK-0013-1	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, qnrS1, sul3, tet_A
EC18PK-0013-2	
EC18PK-0013-3	
EC18PK-0013-4	
EC18PK-0014-1	aac_3__IId, blaTEM, floR, sul2, sul3
EC18PK-0014-2	aac_3__IId, blaTEM, sul2, sul3
EC18PK-0014-3	aac_3__IId, blaTEM, floR, sul2, sul3
EC18PK-0014-4	aac_3__IId, blaTEM, floR, sul2, sul3
EC18PK-0015-2	
EC18PK-0015-3	
EC18PK-0015-4	
EC18PK-0016-1	
EC18PK-0016-2	
EC18PK-0016-3	
EC18PK-0016-4	
EC18PK-0017-1	aadA5, dfrA17, sul2
EC18PK-0017-2	aadA5, dfrA17, sul2
EC18PK-0017-3	
EC18PK-0017-4	
EC18PK-0018-1	
EC18PK-0018-4	parE_I529L
EC18PK-0019-1	blaTEM, tet_A
EC18PK-0019-3	blaTEM, tet_A
EC18PK-0019-4	blaTEM
EC18PK-0020-1	
EC18PK-0021-1	
EC18PK-0021-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0021-3	
EC18PK-0021-4	
EC18PK-0022-1	
EC18PK-0022-2	
EC18PK-0022-3	
EC18PK-0022-4	dfrA5
EC18PK-0023-1	aadA1, sul1, tet_A
EC18PK-0024-1	
EC18PK-0024-2	
EC18PK-0024-3	
EC18PK-0024-4	
EC18PK-0025-1	parC_A56T
EC18PK-0025-2	
EC18PK-0025-3	
EC18PK-0025-4	
EC18PK-0026-3	
EC18PK-0027-1	aph_6__Id, dfrA14, sul2
EC18PK-0027-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, sul2
EC18PK-0027-3	fosA7
EC18PK-0027-4	fosA7
EC18PK-0028-1	
EC18PK-0028-2	
EC18PK-0028-3	
EC18PK-0028-4	
EC18PK-0029-1	
EC18PK-0029-2	
EC18PK-0029-3	
EC18PK-0029-4	
EC18PK-0030-1	
EC18PK-0030-2	tet_B
EC18PK-0030-3	tet_B
EC18PK-0030-4	
EC18PK-0032-1	tet_A
EC18PK-0032-2	tet_A
EC18PK-0032-3	tet_A
EC18PK-0032-4	tet_A
EC18PK-0033-1	
EC18PK-0033-2	
EC18PK-0033-3	
EC18PK-0033-4	
EC18PK-0035-1	
EC18PK-0035-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0035-3	
EC18PK-0035-4	
EC18PK-0036-1	
EC18PK-0036-2	
EC18PK-0036-3	
EC18PK-0036-4	23S r.2032T>C
EC18PK-0038-1	tet_B
EC18PK-0038-2	tet_B
EC18PK-0038-3	tet_B
EC18PK-0038-4	tet_B
EC18PK-0040-1	
EC18PK-0040-2	
EC18PK-0041-1	
EC18PK-0041-2	
EC18PK-0058-1	aph_3____Ib, aph_6__Id, sul2, tet_A
EC18PK-0058-2	blaTEM, tet_A
EC18PK-0059-1	dfrA5, sul1
EC18PK-0059-2	
EC18PK-0064-1	tet_A
EC18PK-0064-2	tet_A
EC18PK-0065-1	aadA1, aadA2, blaTEM, cmlA1, dfrA12, floR, qnrS1, sul2, sul3, tet_A, tet_M
EC18PK-0065-2	
EC18PK-0066-1	
EC18PK-0067-1	blaTEM, tet_A
EC18PK-0067-2	blaTEM, tet_A
EC18PK-0069-1	blaTEM, parE_I529L, tet_A
EC18PK-0069-2	blaTEM, parE_I529L, tet_A
EC18PK-0070-1	
EC18PK-0070-2	
EC18PK-0071-1	
EC18PK-0071-2	
EC18PK-0073-1	tet_A
EC18PK-0073-2	tet_A
EC18PK-0074-1	
EC18PK-0074-2	
EC18PK-0075-1	
EC18PK-0075-2	
EC18PK-0076-1	
EC18PK-0076-2	
EC18PK-0077-1	aadA1, sul1
EC18PK-0077-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0078-1	
EC18PK-0078-2	
EC18PK-0079-1	aadA1, parC_A56T
EC18PK-0079-2	aadA1, parC_A56T
EC18PK-0080-2	
EC18PK-0081-1	
EC18PK-0081-2	
EC18PK-0082-1	aph_3____Ib, aph_6__Id, blaTEM, tet_B
EC18PK-0082-2	aph_3____Ib, aph_6__Id, blaTEM, tet_B
EC18PK-0084-1	aadA1, aph_3____Ib, blaTEM, dfrA1, sul1, sul2, tet_A
EC18PK-0084-2	fosA7, tet_B
EC18PK-0085-1	tet_A
EC18PK-0085-2	tet_A
EC18PK-0086-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, gyrA_D87Y, sul2, tet_A
EC18PK-0087-1	aph_3____Ib, aph_6__Id, fosA7
EC18PK-0087-2	aph_3____Ib, aph_6__Id
EC18PK-0088-1	
EC18PK-0088-2	23S r.2032T>C
EC18PK-0089-1	aadA1, gyrA_S83L, tet_A
EC18PK-0089-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM_99, dfrA1, sul3, tet_B
EC18PK-0090-1	aph_3____Ib, aph_6__Id, blaTEM, sul2, tet_A
EC18PK-0090-2	aph_3____Ib, aph_6__Id, blaTEM, sul2, tet_A
EC18PK-0091-1	aph_3____Ib, aph_6__Id, blaTEM, sul2, tet_A
EC18PK-0091-2	aph_3____Ib, aph_6__Id, blaTEM, sul2, tet_A
EC18PK-0098-1	
EC18PK-0098-2	
EC18PK-0099-1	
EC18PK-0099-2	
EC18PK-0107-1	
EC18PK-0119-2	aadA1, dfrA1
EC18PK-0120-1	
EC18PK-0120-2	aadA1, aph_3____Ib, aph_6__Id, dfrA1, floR, sul2, tet_A
EC18PK-0121-1	
EC18PK-0121-2	
EC18PK-0122-1	aph_3____Ib, aph_3___Ia, aph_6__Id, sul2, tet_B
EC18PK-0122-2	aph_3____Ib, aph_3___Ia, aph_6__Id, sul2, tet_B
EC18PK-0123-1	blaTEM, tet_B
EC18PK-0123-2	aadA1, aadA2, blaTEM, cmlA1, dfrA12, floR, qnrS1, sul3
EC18PK-0124-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, fosA7, sul2
EC18PK-0124-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, fosA7, sul2

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0126-1	
EC18PK-0126-2	
EC18PK-0127-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, parE_I529L, sul2
EC18PK-0127-2	aadA24, aph_3____Ib, aph_6__Id, blaTEM, dfrB1, gyrA_D87Y, tet_B
EC18PK-0128-1	
EC18PK-0128-2	tet_B
EC18PK-0130-1	
EC18PK-0130-2	
EC18PK-0131-1	aph_6__Id, dfrA14, sul2
EC18PK-0131-2	aph_6__Id, dfrA14, sul2
EC18PK-0132-1	aph_6__Id, dfrA14, sul2
EC18PK-0132-2	aph_6__Id, dfrA14, sul2
EC18PK-0133-1	aadA1, tet_A
EC18PK-0133-2	aadA1, blaTEM, dfrA1, sul3, tet_A
EC18PK-0134-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA1, floR, sul2, tet_A
EC18PK-0134-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA1, floR, sul2, tet_A
EC18PK-0135-1	aadA2, blaTEM, cmlA1, dfrA12, tet_B
EC18PK-0135-2	aadA2, aadA8b, blaTEM, cmlA1, dfrA12, tet_B
EC18PK-0136-1	aph_3____Ib, aph_6__Id, blaTEM, catA2, dfrA5, sul2, tet_A
EC18PK-0136-2	aph_3____Ib, aph_6__Id, blaTEM, catA2, dfrA5, sul2, tet_A
EC18PK-0137-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, sul2, tet_A
EC18PK-0137-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM, tet_A
EC18PK-0138-1	aadA1, aadA2, blaTEM, cmlA1, dfrA12, floR, tet_A, tet_M
EC18PK-0140-1	
EC18PK-0140-2	gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A
EC18PK-0141-1	
EC18PK-0141-2	
EC18PK-0142-1	
EC18PK-0142-2	
EC18PK-0143-1	tet_A
EC18PK-0143-2	tet_A
EC18PK-0144-1	blaTEM, parE_I529L, tet_A
EC18PK-0148-1	
EC18PK-0148-2	
EC18PK-0149-1	aadA1, blaTEM, dfrA1, sul1, tet_A
EC18PK-0149-2	fosA7
EC18PK-0153-1	
EC18PK-0153-2	
EC18PK-0154-1	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0154-2	
EC18PK-0203-1	
EC18PK-0203-2	
EC18PK-0203-3	
EC18PK-0220-1	aadA1, blaTEM, dfrA1, qnrB19, sul2
EC18PK-0220-2	aadA1, blaTEM, dfrA1, qnrB19, sul2
EC18PK-0221-1	
EC18PK-0221-2	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, cmlA1, dfrA12, floR, sul2, sul3, tet_A, tet_M
EC18PK-0221-3	aadA1, aadA2, blaTEM, cmlA1, dfrA12, floR, sul2, sul3, tet_A, tet_M
EC18PK-0221-4	blaTEM, dfrA1, sul2, tet_A
EC18PK-0224-1	blaTEM
EC18PK-0224-2	
EC18PK-0225-1	aadA1, tet_B
EC18PK-0225-2	
EC18PK-0225-3	
EC18PK-0225-4	
EC18PK-0230-1	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0230-2	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0230-3	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0230-4	aadA1, aph_3____Ib, aph_6__Id, blaTEM, catA1, dfrA1, sul3, tet_A
EC18PK-0235-1	
EC18PK-0235-2	
EC18PK-0235-3	
EC18PK-0235-4	
EC18PK-0240-1	
EC18PK-0240-2	
EC18PK-0240-3	
EC18PK-0240-4	
EC18PK-0242-1	aadA1, blaTEM, dfrA1, sul1, tet_A
EC18PK-0242-2	
EC18PK-0242-3	aph_6__Id, dfrA14, sul2
EC18PK-0242-4	
EC18PK-0243-1	cmlA1, tet_A
EC18PK-0243-2	cmlA1, tet_A
EC18PK-0250-1	
EC18PK-0250-2	
EC18PK-0251-1	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PK-0251-2	blaTEM, tet_A
EC18PK-0255-1	
EC18PK-0258-1	
EC18PK-0259-1	aadA1, dfrA1, sul1, tet_A
EC18PK-0259-2	aph_3____Ib, aph_6__Id, tet_B
EC18PK-0264-2	
EC18PK-0265-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, sul2, tet_B
EC18PK-0265-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA5, sul2, tet_B
EC18PK-0266-1	
EC18PK-0266-2	
EC18PK-0266-3	cmlA1, tet_A
EC18PK-0267-1	
EC18PK-0267-2	
EC18PK-0272-1	
EC18PK-0272-2	
EC18PK-0272-3	
EC18PK-0275-1	aph_6__Id, dfrA14, sul2, tet_A
EC18PK-0275-2	aph_6__Id, dfrA14, sul2, tet_A
EC18PK-0280-1	
EC18PK-0280-2	parE_I529L
EC18PK-0280-3	parE_I529L
EC18PK-0285-2	
EC18PK-0290-2	tet_B
EC18PK-0292-1	
EC18PK-0306-1	
EC18PK-0306-2	aadA1, blaOXA_1, sul1
EC18PR-0001-1	
EC18PR-0001-2	aph_3____Ib, aph_6__Id, blaCMY_2, blaTEM, dfrA14, floR, qnrS1, sul2
EC18PR-0001-3	
EC18PR-0001-4	
EC18PR-0004-1	blaTEM
EC18PR-0004-2	blaTEM
EC18PR-0004-3	blaTEM
EC18PR-0004-4	
EC18PR-0006-1	
EC18PR-0006-2	
EC18PR-0006-3	
EC18PR-0006-4	
EC18PR-0007-1	blaTEM, floR, tet_M
EC18PR-0007-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PR-0007-3	aadA1, aadA2, blaTEM, dfrA12, gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A, sul3, tet_A
EC18PR-0007-4	
EC18PR-0008-1	
EC18PR-0008-2	
EC18PR-0008-3	
EC18PR-0008-4	
EC18PR-0009-1	
EC18PR-0009-2	
EC18PR-0009-3	
EC18PR-0009-4	
EC18PR-0011-2	blaTEM, gyrA_D87N, gyrA_S83L, parC_E84K, parC_S80I, tet_B
EC18PR-0014-1	
EC18PR-0014-2	
EC18PR-0014-3	
EC18PR-0014-4	
EC18PR-0015-1	
EC18PR-0015-2	
EC18PR-0015-3	
EC18PR-0015-4	
EC18PR-0016-1	
EC18PR-0016-2	
EC18PR-0016-3	
EC18PR-0016-4	
EC18PR-0018-1	
EC18PR-0018-2	
EC18PR-0018-3	
EC18PR-0018-4	
EC18PR-0024-1	
EC18PR-0025-1	
EC18PR-0025-2	
EC18PR-0025-3	
EC18PR-0025-4	
EC18PR-0026-1	
EC18PR-0026-2	
EC18PR-0026-3	
EC18PR-0026-4	
EC18PR-0027-1	
EC18PR-0027-2	
EC18PR-0027-3	
EC18PR-0027-4	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PR-0033-1	
EC18PR-0034-1	aph_6__Id, blaTEM, dfrA14, floR, qnrS1, sul2, tet_A
EC18PR-0034-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A, sul2, tet_A
EC18PR-0034-3	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A, sul2, tet_A
EC18PR-0034-4	aph_6__Id, blaTEM, dfrA14, gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A, sul2
EC18PR-0035-1	blaLAP_2, qnrS1, sul2, tet_A
EC18PR-0035-2	aadA5, dfrA17, sul2, tet_A
EC18PR-0035-3	aadA5, dfrA17, sul2, tet_A
EC18PR-0035-4	aadA5, dfrA17, sul2, tet_A
EC18PR-0037-1	aadA1, aadA2, blaCMY_2, blaTEM, cmlA1, dfrA12, mef_B, sul3, tet_B
EC18PR-0037-2	
EC18PR-0037-3	
EC18PR-0037-4	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, floR, qnrS1, sul2, tet_A
EC18PR-0044-1	
EC18PR-0044-2	
EC18PR-0048-1	tet_A
EC18PR-0048-2	tet_A
EC18PR-0051-1	
EC18PR-0051-2	
EC18PR-0055-1	
EC18PR-0055-2	
EC18PR-0055-3	
EC18PR-0055-4	
EC18PR-0059-1	
EC18PR-0059-2	
EC18PR-0059-3	
EC18PR-0059-4	
EC18PR-0067-1	parC_S57T, parE_I355T
EC18PR-0067-2	
EC18PR-0067-3	parC_S57T, parE_I355T
EC18PR-0067-4	parC_S57T, parE_I355T
EC18PR-0071-1	
EC18PR-0071-2	
EC18PR-0075-1	
EC18PR-0075-2	
EC18PR-0076-1	tet_A
EC18PR-0076-2	aadA5, dfrA17, qnrB7, sul2, tet_A
EC18PR-0078-1	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PR-0078-2	
EC18PR-0078-3	
EC18PR-0078-4	gyrA_S83L
EC18PR-0081-1	
EC18PR-0081-2	
EC18PR-0081-3	
EC18PR-0081-4	
EC18PR-0082-1	blaTEM, tet_A
EC18PR-0082-2	blaTEM, tet_A
EC18PR-0082-3	blaTEM, tet_A
EC18PR-0082-4	blaTEM, tet_A
EC18PR-0086-1	
EC18PR-0086-2	
EC18PR-0086-3	
EC18PR-0086-4	
EC18PR-0090-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, floR, sul2, tet_A
EC18PR-0090-2	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, floR, sul2, tet_A
EC18PR-0094-1	
EC18PR-0094-2	
EC18PR-0094-3	
EC18PR-0094-4	
EC18PR-0097-1	
EC18PR-0097-2	
EC18PR-0097-3	
EC18PR-0097-4	
EC18PR-0107-1	blaTEM, gyrA_D87N, gyrA_S83L, parC_E84K, parC_S80I, tet_B
EC18PR-0107-2	aadA1, aadA2, cmlA1, dfrA12, gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A, sul3, tet_A
EC18PR-0109-1	
EC18PR-0109-2	
EC18PR-0110-1	
EC18PR-0110-2	
EC18PR-0114-1	
EC18PR-0114-2	
EC18PR-0116-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, gyrA_S83L, sul2, tet_A
EC18PR-0116-2	gyrA_S83L
EC18PR-0118-1	
EC18PR-0118-2	
EC18PR-0119-1	
EC18PR-0119-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PR-0123-1	
EC18PR-0123-2	
EC18PR-0125-1	
EC18PR-0125-2	
EC18PR-0126-1	
EC18PR-0126-2	
EC18PR-0131-1	dfrA15, qnrS1, tet_A
EC18PR-0131-2	
EC18PR-0132-1	
EC18PR-0132-2	
EC18PR-0133-1	
EC18PR-0137-1	
EC18PR-0137-2	
EC18PR-0143-1	
EC18PR-0143-2	
EC18PR-0150-1	
EC18PR-0150-2	
EC18PR-0151-1	aph_3____Ib, aph_6__Id, blaTEM, dfrA14, floR, qnrS1, sul2, tet_A
EC18PR-0151-2	
EC18PR-0154-1	
EC18PR-0154-2	
EC18PR-0157-1	
EC18PR-0157-2	
EC18PR-0160-1	
EC18PR-0160-2	23S r.754G>A
EC18PR-0163-1	
EC18PR-0163-2	
EC18PR-0164-1	
EC18PR-0164-2	
EC18PR-0166-1	
EC18PR-0166-2	
EC18PR-0170-1	
EC18PR-0170-2	
EC18PR-0172-1	sul1, tet_B
EC18PR-0172-2	sul1, tet_B
EC18PR-0173-1	gyrA_S83A
EC18PR-0173-2	gyrA_S83A
EC18PR-0174-1	gyrA_S83A
EC18PR-0174-2	dfrA14, qnrS1, sul2, tet_A
EC18PR-0175-1	
EC18PR-0175-2	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18PR-0177-1	
EC18PR-0177-2	
EC18PR-0178-1	parE_I355T
EC18PR-0178-2	parE_I355T
EC18PR-0179-1	
EC18PR-0179-2	blaTEM, tet_A
EC18PR-0183-1	aadA1, aadA2, aph_3____Ib, aph_6__Id, blaTEM, cmlA1, dfrA12, floR, qnrS1, sul2, sul3
EC18PR-0183-2	blaCTX_M_55, blaTEM, mph_A, qnrS13, tet_A
EC18PR-0184-1	
EC18PR-0184-2	
EC18PR-0189-1	qnrS1
EC18PR-0189-2	qnrS1
EC18PR-0189-3	qnrS1
EC18PR-0189-4	qnrS1
EC18PR-0190-1	
EC18PR-0190-2	
EC18PR-0190-3	
EC18PR-0190-4	aadA1, aadA2, cmlA1, dfrA12, sul3, tet_A
EC18PR-0191-1	aadA5, dfrA17, sul2, tet_A
EC18PR-0191-2	aadA5, dfrA17, sul2, tet_A
EC18PR-0191-3	
EC18PR-0191-4	aadA5, dfrA17, sul2, tet_A
EC18PR-0197-1	parE_I355T
EC18PR-0197-2	gyrA_S83L, tet_A
EC18PR-0197-3	parE_I355T
EC18PR-0197-4	parE_I355T
EC18PR-0251-1	
EC18PR-0251-2	
EC18PR-0251-3	
EC18PR-0251-4	
EC18PR-0256-1	aadA1, blaTEM, dfrA1, qnrS4, tet_A
EC18PR-0256-2	aadA1, blaTEM, dfrA1, qnrS4, tet_A
EC18PR-0256-3	aadA1, blaTEM, dfrA1, qnrS4, tet_A
EC18PR-0262-1	sul1, tet_B
EC18SM-0006-1	
EC18SM-0006-2	
EC18SM-0006-3	
EC18SM-0006-4	
EC18SM-0007-1	aadA1, aadA2, blaCARB_2, cmlA1, parE_I355T, sul3, tet_A
EC18SM-0007-2	parE_I355T

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18SM-0007-3	parE_I355T
EC18SM-0007-4	parE_I355T
EC18SM-0013-1	
EC18SM-0013-2	
EC18SM-0013-3	
EC18SM-0013-4	
EC18SM-0015-1	
EC18SM-0015-2	
EC18SM-0015-3	
EC18SM-0015-4	
EC18SM-0021-1	
EC18SM-0021-2	
EC18SM-0021-3	
EC18SM-0021-4	
EC18SM-0028-1	parC_S57T
EC18SM-0028-2	parC_S57T
EC18SM-0029-1	
EC18SM-0029-2	
EC18SM-0042-1	
EC18SM-0049-1	
EC18SM-0049-2	
EC18SM-0054-1	
EC18SM-0054-2	
EC18SM-0054-3	
EC18SM-0054-4	
EC18SM-0059-1	
EC18SM-0059-2	
EC18SM-0060-1	
EC18SM-0060-2	
EC18SM-0067-1	
EC18SM-0067-2	
EC18SM-0067-3	
EC18SM-0067-4	
EC18SM-0069-1	
EC18SM-0069-2	
EC18SM-0069-3	
EC18SM-0069-4	
EC18SM-0075-1	
EC18SM-0075-2	
EC18SM-0075-3	
EC18SM-0075-4	
EC18SM-0078-1	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18SM-0078-2	
EC18SM-0078-3	
EC18SM-0078-4	23S r.2058T>G
EC18SM-0082-1	
EC18SM-0082-2	
EC18SM-0082-3	
EC18SM-0082-4	
EC18SM-0084-1	
EC18SM-0088-1	aadA1, ARR_2, blaOXA_10, cmlA1, dfrA14, floR, qnrS1
EC18SM-0088-2	aadA1, ARR_2, blaOXA_10, cmlA1, dfrA14, floR, qnrS1
EC18SM-0105-1	parC_S57T
EC18SM-0105-2	
EC18SM-0105-3	
EC18SM-0105-4	
EC18SM-0112-1	
EC18SM-0112-2	
EC18SM-0112-3	
EC18SM-0112-4	
EC18SM-0116-1	23S r.2032T>C
EC18SM-0116-2	
EC18SM-0116-3	23S r.2032T>C
EC18SM-0116-4	23S r.2032T>C
EC18SM-0119-1	23S r.2032T>C
EC18SM-0119-2	23S r.2032T>C
EC18SM-0119-3	
EC18SM-0119-4	
EC18SM-0120-1	
EC18SM-0120-2	
EC18SM-0120-3	
EC18SM-0120-4	
EC18SM-0121-1	
EC18SM-0121-2	
EC18SM-0121-3	
EC18SM-0121-4	
EC18SM-0124-1	
EC18SM-0124-2	
EC18SM-0124-3	
EC18SM-0125-1	
EC18SM-0125-2	
EC18SM-0125-3	
EC18SM-0125-4	
EC18SM-0127-1	

Appendix 4: Antimicrobial resistance determinants identified in the 1,067 *Escherichia coli* genomes using ARIBA and PointFinder.

Genome	AMR determinant genotype
EC18SM-0127-2	
EC18SM-0127-3	23S r.2032T>C
EC18SM-0127-4	
EC18SM-0130-1	
EC18SM-0130-2	
EC18SM-0130-3	
EC18SM-0130-4	
EC18SM-0135-1	
EC18SM-0135-2	
EC18SM-0135-3	
EC18SM-0137-1	
EC18SM-0137-2	
EC18SM-0149-1	blaTEM, tet_A
EC18SM-0149-2	blaTEM, tet_A
EC18SM-0149-3	blaTEM, tet_A
EC18SM-0149-4	blaTEM, tet_A
EC18SM-0152-1	
EC18SM-0152-2	
EC18SM-0152-3	
EC18SM-0152-4	

Appendix 5: List of antimicrobial resistance genes and associated drug classifications.

ARIBA AMR code	Gene ID	NCBI accession	Drug class summary	Chapter
aac_3__IId.1_EU022314_2	aac_3__IId	EU022314	Aminoglycoside	Chapter 4,5
aac_3__IV.1_DQ241380_2	aac_3__IV	DQ241380	Aminoglycoside	Chapter 4
aac_3__VIa.2_NC_009838_2	aac_3__VIa	NC_009838	Aminoglycoside	Chapter 4
aadA1.2_FJ591054_2	aadA1	FJ591054	Aminoglycoside	Chapter 4
aadA1.3_JQ414041_2	aadA1	JQ414041	Aminoglycoside	Chapter 4,5
aadA1.4_JQ480156_2	aadA1	JQ480156	Aminoglycoside	Chapter 4
aadA1.5_JX185132_2	aadA1	JX185132	Aminoglycoside	Chapter 4,5
aadA17.1_FJ460181_2	aadA17	FJ460181	Aminoglycoside	Chapter 4
aadA2.2_JQ364967_2	aadA2	JQ364967	Aminoglycoside	Chapter 4,5
aadA22.1_AM261837_2	aadA22	AM261837	Aminoglycoside	Chapter 4
aadA24.1_DQ677333_2	aadA24	DQ677333	Aminoglycoside	Chapter 4
aadA2b.1_D43625_2	aadA2b	D43625	Aminoglycoside	Chapter 4
aadA5.1_AF137361_2	aadA5	AF137361	Aminoglycoside	Chapter 4
aadA8b.2_AM040708_2	aadA8b	AM040708	Aminoglycoside	Chapter 4
aph_3____Ib.5_AF321551_2	aph_3____Ib	AF321551	Aminoglycoside	Chapter 4,5
aph_3___Ia.1_V00359_2	aph_3___Ia	V00359	Aminoglycoside	Chapter 4
aph_3___Ia.9_EU722351_2	aph_3___Ia	EU722351	Aminoglycoside	Chapter 4
aph_4__Ia.1_V01499_2	aph_4__Ia	V01499	Aminoglycoside	Chapter 4
aph_6__Id.1_M28829_2	aph_6__Id	M28829	Aminoglycoside	Chapter 4,5
aph_6__Id.4_CP000971_2	aph_6__Id	CP000971	Aminoglycoside	Chapter 4,5
ARR_2.1_HQ141279_2	ARR_2	HQ141279	Rifamycin	Chapter 4
blaCARB_2.1_M69058_2	blaCARB_2	M69058	Beta-lactam	Chapter 4
blaCMY_2.1_X91840_2	blaCMY_2	X91840	Beta-lactam	Chapter 4,5
blaCTX_M_15.1_AY044436_2	blaCTX_M_15	AY044436	Beta-lactam	Chapter 5
blaCTX_M_27.1_AY156923_2	blaCTX_M_27	AY156923	Beta-lactam	Chapter 4
blaCTX_M_55.1_DQ810789_2	blaCTX_M_55	DQ810789	Beta-lactam	Chapter 4
blaLAP_2.1_EU159120_2	blaLAP_2	EU159120	Beta-lactam	Chapter 4

Appendix 5: List of antimicrobial resistance genes and associated drug classifications.

ARIBA AMR code	Gene ID	NCBI accession	Drug class summary	Chapter
blaOXA_1.1_HQ170510_2	blaOXA_1	HQ170510	Beta-lactam	Chapter 4
blaOXA_10.1_J03427_2	blaOXA_10	J03427	Beta-lactam	Chapter 4
blaSHV_48.1_AY263404_2	blaSHV_48	AY263404	Beta-lactam	Chapter 5
blaTEM_1A.1_HM749966_2	blaTEM_1A	HM749966	Beta-lactam	Chapter 4,5
blaTEM_1B.1_AY458016_2	blaTEM_1B	AY458016	Beta-lactam	Chapter 4,5
blaTEM_1C.1_FJ560503_2	blaTEM_1C	FJ560503	Beta-lactam	Chapter 4
blaTEM_1D.1_AF188200_2	blaTEM_1D	AF188200	Beta-lactam	Chapter 4
blaTEM_220.1_KM998962_2	blaTEM_220	KM998962	Beta-lactam	Chapter 4
blaTEM_30.1_AJ437107_2	blaTEM_30	AJ437107	Beta-lactam	Chapter 4
blaTEM_40.1_FR717535_2	blaTEM_40	FR717535	Beta-lactam	Chapter 4
blaTEM_99.1_AF397066_2	blaTEM_99	AF397066	Beta-lactam	Chapter 4
catA1.1_V00622_2	catA1	V00622	Chloramphenicol	Chapter 4
catA2.1_X53796_2	catA2	X53796	Chloramphenicol	Chapter 4
catB3.2_U13880_2	catB3	U13880	Chloramphenicol	Chapter 4,5
cmlA1.1_M64556_2	cmlA1	M64556	Chloramphenicol	Chapter 4,5
dfrA1.8_X00926_2	dfrA1	X00926	Trimethoprim	Chapter 4,5
dfrA1.9_AJ238350_2	dfrA1	AJ238350	Trimethoprim	Chapter 4,5
dfrA12.8_AM040708_2	dfrA12	AM040708	Trimethoprim	Chapter 4,5
dfrA14.1_KF921535_2	dfrA14	KF921535	Trimethoprim	Chapter 4
dfrA14.4_AF393510_2	dfrA14	AF393510	Trimethoprim	Chapter 4,5
dfrA15.2_AF221900_2	dfrA15	AF221900	Trimethoprim	Chapter 4
dfrA16.2_AF174129_2	dfrA16	AF174129	Trimethoprim	Chapter 4
dfrA17.1_FJ460238_2	dfrA17	FJ460238	Trimethoprim	Chapter 4,5
dfrA36.1_CP038791_2	dfrA36	CP038791	Trimethoprim	Chapter 4
dfrA5.1_X12868_2	dfrA5	X12868	Trimethoprim	Chapter 4
dfrA7.1_AB161450_2	dfrA7	AB161450	Trimethoprim	Chapter 4,5
dfrA8.1_U10186_2	dfrA8	U10186	Trimethoprim	Chapter 4,5

Appendix 5: List of antimicrobial resistance genes and associated drug classifications.

ARIBA AMR code	Gene ID	NCBI accession	Drug class summary	Chapter
dfrB1.1_U36276_2	dfrB1	U36276	Trimethoprim	Chapter 4
dfrB4.1_AJ429132_2	dfrB4	AJ429132	Trimethoprim	Chapter 4
erm_42_.1_FR734406_2	erm_42	FR734406	Multidrug	Chapter 4
erm_B_.1_JN899585_2	erm_B	JN899585	Multidrug	Chapter 5
floR.2_AF118107_2	floR	AF118107	Chloramphenicol	Chapter 4,5
fosA7.1_LAPJ01000014_2	fosA7	LAPJ01000014	Fosfomycin	Chapter 4
lnu_F_.1_EU118119_2	lnu_F	EU118119	Lincosamide	Chapter 4
lnu_G_.1_KX470419_2	lnu_G	KX470419	Lincosamide	Chapter 4
mcr_1.26.1_NG_068217_2	mcr_1	NG_068217	Colistin	Chapter 4
mef_B_.1_FJ196385_2	mef_B	FJ196385	Macrolide	Chapter 4
mph_A_.1_D16251_2	mph_A	D16251	Macrolide	Chapter 4,5
mph_B_.1_D85892_2	mph_B	D85892	Macrolide	Chapter 4
qnrB19.1_EU432277_2	qnrB19	EU432277	Quinolone	Chapter 4,5
qnrB7.1_EU043311_2	qnrB7	EU043311	Quinolone	Chapter 4
qnrS1.1_AB187515_2	qnrS1	AB187515	Quinolone	Chapter 4,5
qnrS13.1_LUYD01000008_2	qnrS13	LUYD01000008	Quinolone	Chapter 4
qnrS4.1_FJ418153_2	qnrS4	FJ418153	Quinolone	Chapter 4
sul1.2_U12338_2	sul1	U12338	Sulphonamide	Chapter 4,5
sul1.39_AY522923_2	sul1	AY522923	Sulphonamide	Chapter 4
sul1.9_AY963803_2	sul1	AY963803	Sulphonamide	Chapter 4
sul2.2_AY034138_2	sul2	AY034138	Sulphonamide	Chapter 4,5
sul2.3_HQ840942_2	sul2	HQ840942	Sulphonamide	Chapter 4
sul2.6_FN995456_2	sul2	FN995456	Sulphonamide	Chapter 4,5
sul3.2_AJ459418_2	sul3	AJ459418	Sulphonamide	Chapter 4,5
tet_A_.4_AJ517790_2	tet_A	AJ517790	Tetracycline	Chapter 4,5
tet_A_.6_AF534183_2	tet_A	AF534183	Tetracycline	Chapter 4
tet_B_.1_AP000342_2	tet_B	AP000342	Tetracycline	Chapter 4

Appendix 5: List of antimicrobial resistance genes and associated drug classifications.

ARIBA AMR code	**Gene ID**	**NCBI accession**	**Drug class summary**	**Chapter**
tet_B_.2_AF326777_2	tet_B	AF326777	Tetracycline	Chapter 4,5
tet_C_.3_AF055345_2	tet_C	AF055345	Tetracycline	Chapter 4
tet_M_.5_U58985_2	tet_M	U58985	Tetracycline	Chapter 4
tet_M_.8_X04388_2	tet_M	X04388	Tetracycline	Chapter 4

Appendix 6: List of point mutations conferring AMR and associated drug classifications.

Point mutation	Amino acid change	PMID number	Drug class summary	Chapter
23S r.2032G>C	G -> C	10986233	Oxazolidinone	Chapter 5
23S r.2032T>C	T -> C	10986233	Oxazolidinone	Chapter 4, 5
23S r.2058T>G	T -> G	15616307	Macrolide	Chapter 4
23S r.754G>A	G -> A	10027979	Macrolide	Chapter 4
gyrA p.D87G	GAC -> GGC	12654733	Quinolone	Chapter 5
gyrA p.D87N	GAC -> AAC	12654733	Quinolone	Chapter 4, 5
gyrA p.S83L	TCG -> TTG	8891148	Quinolone	Chapter 4, 5
parC p.E84K	GAA -> AAA	8524852	Quinolone	Chapter 4
parC p.S57T	AGC -> ACC	14510643	Quinolone	Chapter 4, 5
parC p.S80I	AGC -> ATC	8851598	Quinolone	Chapter 4, 5
parE p.I355T	ATC -> ACC	28598203	Quinolone	Chapter 4, 5
parE p.I529L	ATT -> CTT	14506034	Quinolone	Chapter 5
parE p.L416F	CTT -> TTT	14510643	Quinolone	Chapter 5
parE p.S458T	TCG -> ACG	14506034	Quinolone	Chapter 4

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23D1-001-1	Dog	D1	001	EC23D1-001	961	B2	O4:H5
EC23D1-001-2	Dog	D1	001	EC23D1-001	961	B2	O4:H5
EC23D1-001-2	Dog	D1	001	EC23D1-001	961	B2	O4:H5
EC23D1-003-1	Dog	D1	003	EC23D1-003	1442	B2	O8:H10
EC23D1-003-2	Dog	D1	003	EC23D1-003	978	B2	O4:H27
EC23D1-006-1	Dog	D1	006	EC23D1-006	3576	B1	O8:H7
EC23D1-006-2	Dog	D1	006	EC23D1-006	3576	B1	O8:H7
EC23D1-007-1	Dog	D1	007	EC23D1-007	642	B1	O85:H4
EC23D1-007-2	Dog	D1	007	EC23D1-007	novel6	B1	O180:H12
EC23D1-009-1	Dog	D1	009	EC23D1-009	8573	D	O8:-
EC23D1-009-2	Dog	D1	009	EC23D1-009	8573	D	O8:-
EC23D1-010-1	Dog	D1	010	EC23D1-010	2163	B1	-:H37
EC23D1-010-2	Dog	D1	010	EC23D1-010	2163	B1	-:H37
EC23D1-011-1	Dog	D1	011	EC23D1-011	101	B1	O102:H6
EC23D1-012-1	Dog	D1	012	EC23D1-012	2538	D	O167:H9
EC23D1-012-2	Dog	D1	012	EC23D1-012	589	B2	O51:H49
EC23D1-013-1	Dog	D1	013	EC23D1-013	399	A	O27:H7
EC23D1-013-2	Dog	D1	013	EC23D1-013	399	A	O27:H7
EC23D1-016-1	Dog	D1	016	EC23D1-016	12	B2	O4:H5
EC23D1-016-2	Dog	D1	016	EC23D1-016	12	B2	O4:H5
EC23D1-023-1	Dog	D1	023	EC23D1-023	6164	B2	O175:H5
EC23D1-026-1	Dog	D1	026	EC23D1-026	101	B1	O93:H8
EC23D1-026-2	Dog	D1	026	EC23D1-026	101	B1	O93:H8

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23D1-030-1	Dog	D1	030	EC23D1-030	7345	cladeV	-:H56
EC23D1-030-2	Dog	D1	030	EC23D1-030	7345	cladeV	-:H56
EC23D1-031-1	Dog	D1	031	EC23D1-031	10	A	O26:H32
EC23D1-031-2	Dog	D1	031	EC23D1-031	10	A	O26:H32
EC23D1-033-1	Dog	D1	033	EC23D1-033	12	B2	O4:H5
EC23D1-033-2	Dog	D1	033	EC23D1-033	12	B2	O4:H5
EC23D1-038-1	Dog	D1	038	EC23D1-038	106	D	O17/O44/O73/O77/O106:H18
EC23D1-038-2	Dog	D1	038	EC23D1-038	106	D	O17/O44/O77/O106:H18
EC23D1-045-1	Dog	D1	045	EC23D1-045	7236	cladeV	-:H56
EC23D1-045-2	Dog	D1	045	EC23D1-045	7236	cladeV	-:H56
EC23D1-046-1	Dog	D1	046	EC23D1-046	942	B1	O163:H28
EC23D1-047-1	Dog	D1	047	EC23D1-047	101	B1	O142:H38
EC23D1-047-2	Dog	D1	047	EC23D1-047	101	B1	O142:H38
EC23D1-048-1	Dog	D1	048	EC23D1-048	131	B2	O16:H5
EC23D1-048-2	Dog	D1	048	EC23D1-048	131	B2	O16:H5
EC23D1-049-1	Dog	D1	049	EC23D1-049	1800	B1	O82:H21
EC23D1-049-2	Dog	D1	049	EC23D1-049	1800	B1	O82:H21

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23D1-053-1	Dog	D1	053	EC23D1-053	5415	B1	O153/O178:H7
EC23D1-053-2	Dog	D1	053	EC23D1-053	5415	B1	O153/O178:H7
EC23D1-056-1	Dog	D1	056	EC23D1-056	131	B2	O16:H5
EC23D1-056-2	Dog	D1	056	EC23D1-056	131	B2	O16:H5
EC23D2-038-1	Dog	D2	038	EC23D2-038	2559	cladeV	-:H56
EC23D2-038-2	Dog	D2	038	EC23D2-038	101	B1	O93:H8
EC23D3-044-1	Dog	D3	044	EC23D3-044	1079	B1	O6:H49
EC23D3-044-2	Dog	D3	044	EC23D3-044	164	B1	O70:H42
EC23H1-001-1	Human	H1	001	EC23H1-001	8125	A	O112:H19
EC23H1-001-2	Human	H1	001	EC23H1-001	8125	A	O112:H19
EC23H1-003-1	Human	H1	003	EC23H1-003	69	D	O15:H1
EC23H1-003-2	Human	H1	003	EC23H1-003	69	D	O15:H1
EC23H1-006-1	Human	H1	006	EC23H1-006	141	B2	O2:H6
EC23H1-006-2	Human	H1	006	EC23H1-006	141	B2	O2:H6
EC23H1-007-1	Human	H1	007	EC23H1-007	141	B2	O2:H6
EC23H1-007-2	Human	H1	007	EC23H1-007	141	B2	O2:H6
EC23H1-009-1	Human	H1	009	EC23H1-009	38	D	O153:H30
EC23H1-009-2	Human	H1	009	EC23H1-009	38	D	O153:H30
EC23H1-010-1	Human	H1	010	EC23H1-010	59	F	O1:H7
EC23H1-010-2	Human	H1	010	EC23H1-010	4429	A	O161:H4
EC23H1-011-1	Human	H1	011	EC23H1-011	906	B1	O8:H23

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23H1-011-2	Human	H1	011	EC23H1-011	1079	B1	O6:H49
EC23H1-012-1	Human	H1	012	EC23H1-012	710	A	-:H30
EC23H1-012-2	Human	H1	012	EC23H1-012	710	A	-:H30
EC23H1-013-1	Human	H1	013	EC23H1-013	95	B2	O1:H7
EC23H1-013-2	Human	H1	013	EC23H1-013	95	B2	O1:H7
EC23H1-016-1	Human	H1	016	EC23H1-016	9265	D	O17/ O44/ O73/ O77/ O106:H18
EC23H1-016-2	Human	H1	016	EC23H1-016	9265	D	O17/ O44/ O73/ O77/ O106:H18
EC23H1-023-1	Human	H1	023	EC23H1-023	10	A	O160:H21
EC23H1-023-2	Human	H1	023	EC23H1-023	10	A	O160:H21
EC23H1-026-1	Human	H1	026	EC23H1-026	131	B2	-:H5
EC23H1-026-2	Human	H1	026	EC23H1-026	131	B2	-:H5
EC23H1-030-1	Human	H1	030	EC23H1-030	372	B2	O83:H31
EC23H1-030-2	Human	H1	030	EC23H1-030	404	B2	O75:H5
EC23H1-031-1	Human	H1	031	EC23H1-031	5295	A	O71:H12
EC23H1-031-2	Human	H1	031	EC23H1-031	5295	A	O71:H12
EC23H1-033-1	Human	H1	033	EC23H1-033	5345	cladeIII	O13/

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
							O129/ O135:-
EC23H1-033-2	Human	H1	033	EC23H1-033	5345	cladeIII	O13/ O129/ O135:-
EC23H1-038-1	Human	H1	038	EC23H1-038	10	A	O9:H9
EC23H1-038-2	Human	H1	038	EC23H1-038	70	D	O7:H15
EC23H1-044-1	Human	H1	044	EC23H1-044	542	A	O65:H26
EC23H1-044-2	Human	H1	044	EC23H1-044	131	B2	O16:H5
EC23H1-045-1	Human	H1	045	EC23H1-045	131	B2	O25:H4
EC23H1-045-2	Human	H1	045	EC23H1-045	131	B2	O25:H4
EC23H1-046-1	Human	H1	046	EC23H1-046	34	A	-:H33
EC23H1-046-2	Human	H1	046	EC23H1-046	34	A	-:H33
EC23H1-047-1	Human	H1	047	EC23H1-047	novel1	Non Escherichia	
EC23H1-047-2	Human	H1	047	EC23H1-047	131	B2	O25:H4
EC23H1-048-1	Human	H1	048	EC23H1-048	1193	B2	O75:H5
EC23H1-048-2	Human	H1	048	EC23H1-048	131	B2	O16:H5
EC23H1-049-1	Human	H1	049	EC23H1-049	73	B2	O22:H1
EC23H1-049-2	Human	H1	049	EC23H1-049	73	B2	O22:H1
EC23H1-053-1	Human	H1	053	EC23H1-053	58	B1	-:H37
EC23H1-053-2	Human	H1	053	EC23H1-053	58	B1	-:H37
EC23H2-003-1	Human	H2	003	EC23H2-003	3759	B1	O8:H49

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23H2-003-2	Human	H2	003	EC23H2-003	3759	B1	O8:H49
EC23H2-006-1	Human	H2	006	EC23H2-006	141	B2	O2:H6
EC23H2-006-2	Human	H2	006	EC23H2-006	635	A	O1:H31
EC23H2-010-1	Human	H2	010	EC23H2-010	362	D	O15:H1
EC23H2-010-2	Human	H2	010	EC23H2-010	novel2	Non Escherichia	
EC23H2-016-1	Human	H2	016	EC23H2-016	12	B2	O4:H5
EC23H2-016-2	Human	H2	016	EC23H2-016	12	B2	O4:H5
EC23H2-046-1	Human	H2	046	EC23H2-046	69	D	O15:H6
EC23H2-046-2	Human	H2	046	EC23H2-046	69	D	O15:H6
EC23H2-048-1	Human	H2	048	EC23H2-048	1193	B2	O75:H5
EC23H2-048-2	Human	H2	048	EC23H2-048	1193	B2	O75:H5
EC23H3-045-1	Human	H3	045	EC23H3-045	101	B1	O93:H8
EC23H3-045-2	Human	H3	045	EC23H3-045	1800	B1	O82:H21
EC23D1-001-3	Dog	D1	001	EC23D1-001	961	B2	O4:H5
EC23D1-001-4	Dog	D1	001	EC23D1-001	961	B2	O4:H5
EC23D1-003-3	Dog	D1	003	EC23D1-003	1442	B2	O8:H10
EC23D1-006-3	Dog	D1	006	EC23D1-006	3576	B1	O8:H7
EC23D1-006-4	Dog	D1	006	EC23D1-006	3576	B1	O8:H7
EC23D1-007-3	Dog	D1	007	EC23D1-007	novel6	B1	O180:H12
EC23D1-007-4	Dog	D1	007	EC23D1-007	novel6	B1	O180:H12
EC23D1-007-5	Dog	D1	007	EC23D1-007	642	B1	O85:H4
EC23D1-009-3	Dog	D1	009	EC23D1-009	720	D	O17/

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
							O44/O73/O77/O106:H41
EC23D1-009-4	Dog	D1	009	EC23D1-009	720	D	O17/O44/O77/O106:H41
EC23D1-012-3	Dog	D1	012	EC23D1-012	5254	B2	O131:H21
EC23D1-012-4	Dog	D1	012	EC23D1-012	589	B2	O51:H49
EC23D1-013-3	Dog	D1	013	EC23D1-013	7075	A	O11:H12
EC23D1-013-4	Dog	D1	013	EC23D1-013	399	A	O27:H7
EC23D1-013-5	Dog	D1	013	EC23D1-013	7075	A	O11:H12
EC23D1-013-6	Dog	D1	013	EC23D1-013	7075	A	O11:H12
EC23D1-016-3	Dog	D1	016	EC23D1-016	12	B2	O4:H5
EC23D1-016-4	Dog	D1	016	EC23D1-016	12	B2	O4:H5
EC23D1-026-3	Dog	D1	026	EC23D1-026	349	D	O166:H15
EC23D1-030-4	Dog	D1	030	EC23D1-030	7347	cladeV	-:H56
EC23D1-031-3	Dog	D1	031	EC23D1-031	10	A	O26:H32
EC23D1-031-4	Dog	D1	031	EC23D1-031	10	A	O26:H32
EC23D1-033-3	Dog	D1	033	EC23D1-033	12	B2	O4:H5
EC23D1-038-3	Dog	D1	038	EC23D1-038	106	D	O17/O44/O73/

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
							O77/ O106:H18
EC23D1-038-4	Dog	D1	038	EC23D1-038	106	D	O17/ O44/ O73/ O77/ O106:H18
EC23D1-045-3	Dog	D1	045	EC23D1-045	7236	cladeV	-:H56
EC23D1-045-4	Dog	D1	045	EC23D1-045	7236	cladeV	-:H56
EC23D1-046-4	Dog	D1	046	EC23D1-046	131	B2	O25:H4
EC23D1-047-3	Dog	D1	047	EC23D1-047	101	B1	O142:H38
EC23D1-047-4	Dog	D1	047	EC23D1-047	101	B1	O142:H38
EC23D1-049-3	Dog	D1	049	EC23D1-049	1800	B1	O82:H21
EC23D1-053-3	Dog	D1	053	EC23D1-053	5415	B1	O153/ O178:H7
EC23D1-053-4	Dog	D1	053	EC23D1-053	5415	B1	O153/ O178:H7
EC23D1-056-3	Dog	D1	056	EC23D1-056	372	B2	O4:H31
EC23D1-056-4	Dog	D1	056	EC23D1-056	372	B2	O4:H31
EC23D2-038-4	Dog	D2	038	EC23D2-038	2559	cladeV	-:H56
EC23D3-044-3	Dog	D3	044	EC23D3-044	164	B1	O70:H42
EC23D3-044-4	Dog	D3	044	EC23D3-044	1079	B1	O6:H49
EC23H1-001-3	Human	H1	001	EC23H1-001	8125	A	O112:H19
EC23H1-001-4	Human	H1	001	EC23H1-001	8125	A	O112:H19

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23H1-003-3	Human	H1	003	EC23H1-003	69	D	O15:H1
EC23H1-003-4	Human	H1	003	EC23H1-003	69	D	O15:H1
EC23H1-006-3	Human	H1	006	EC23H1-006	141	B2	O2:H6
EC23H1-006-4	Human	H1	006	EC23H1-006	141	B2	O2:H6
EC23H1-007-3	Human	H1	007	EC23H1-007	141	B2	O2:H6
EC23H1-007-4	Human	H1	007	EC23H1-007	141	B2	O2:H6
EC23H1-009-3	Human	H1	009	EC23H1-009	38	D	O153:H30
EC23H1-009-4	Human	H1	009	EC23H1-009	38	D	O153:H30
EC23H1-010-3	Human	H1	010	EC23H1-010	59	F	O1:H7
EC23H1-010-4	Human	H1	010	EC23H1-010	4429	A	O161:H4
EC23H1-011-3	Human	H1	011	EC23H1-011	906	B1	O8:H23
EC23H1-011-4	Human	H1	011	EC23H1-011	906	B1	O150:H8
EC23H1-012-4	Human	H1	012	EC23H1-012	710	A	-:H30
EC23H1-012-6	Human	H1	012	EC23H1-012	710	A	-:H30
EC23H1-013-3	Human	H1	013	EC23H1-013	95	B2	O1:H7
EC23H1-013-4	Human	H1	013	EC23H1-013	95	B2	O1:H7
EC23H1-013-5	Human	H1	013	EC23H1-013	95	B2	O1:H7
EC23H1-013-6	Human	H1	013	EC23H1-013	95	B2	O1:H7
EC23H1-016-3	Human	H1	016	EC23H1-016	9265	D	O17/O44/O73/O77/O106:H18

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23H1-016-4	Human	H1	016	EC23H1-016	9265	D	O17/O44/O73/O77/O106:H18
EC23H1-023-3	Human	H1	023	EC23H1-023	372	B2	O75:H31
EC23H1-023-4	Human	H1	023	EC23H1-023	novel57	Non Escherichia	
EC23H1-026-3	Human	H1	026	EC23H1-026	131	B2	-:H5
EC23H1-026-4	Human	H1	026	EC23H1-026	131	B2	-:H5
EC23H1-026-5	Human	H1	026	EC23H1-026	131	B2	-:H5
EC23H1-026-6	Human	H1	026	EC23H1-026	131	B2	-:H5
EC23H1-03-5	Human	H1	003	EC23H1-03	69	D	O15:H1
EC23H1-031-3	Human	H1	031	EC23H1-031	5295	A	O71:H12
EC23H1-031-4	Human	H1	031	EC23H1-031	5295	A	O71:H12
EC23H1-031-5	Human	H1	031	EC23H1-031	5295	A	O71:H12
EC23H1-033-3	Human	H1	033	EC23H1-033	95	B2	O2/O50:H4
EC23H1-033-4	Human	H1	033	EC23H1-033	95	B2	O2/O50:H4
EC23H1-038-3	Human	H1	038	EC23H1-038	70	D	O7:H15
EC23H1-038-4	Human	H1	038	EC23H1-038	70	D	O7:H15
EC23H1-044-3	Human	H1	044	EC23H1-044	164	B1	O70:H42
EC23H1-044-4	Human	H1	044	EC23H1-044	2163	B1	-:H37
EC23H1-044-5	Human	H1	044	EC23H1-044	542	A	O65:H26

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23H1-045-3	Human	H1	045	EC23H1-045	131	B2	O25:H4
EC23H1-045-4	Human	H1	045	EC23H1-045	131	B2	O25:H4
EC23H1-046-4	Human	H1	046	EC23H1-046	34	A	-:H33
EC23H1-046-5	Human	H1	046	EC23H1-046	34	A	-:H33
EC23H1-047-3	Human	H1	047	EC23H1-047	131	B2	O25:H4
EC23H1-047-4	Human	H1	047	EC23H1-047	131	B2	O25:H4
EC23H1-049-3	Human	H1	049	EC23H1-049	73	B2	O22:H1
EC23H1-053-3	Human	H1	053	EC23H1-053	58	B1	-:H37
EC23H2-003-3	Human	H2	003	EC23H2-003	3759	B1	O8:H49
EC23H2-003-4	Human	H2	003	EC23H2-003	3759	B1	O8:H49
EC23H2-006-3	Human	H2	006	EC23H2-006	141	B2	O2:H6
EC23H2-006-4	Human	H2	006	EC23H2-006	635	A	O1:H31
EC23H2-016-3	Human	H2	016	EC23H2-016	73	B2	O6:H1
EC23H2-016-4	Human	H2	016	EC23H2-016	12	B2	O4:H5
EC23H2-016-5	Human	H2	016	EC23H2-016	73	B2	O6:H1
EC23H2-048-3	Human	H2	048	EC23H2-048	1193	B2	O75:H5
EC23H2-048-4	Human	H2	048	EC23H2-048	1193	B2	O75:H5
EC24D1-003-4	Dog	D1	003	EC24D1-003	978	B2	O4:H27
EC24D1-010-3	Dog	D1	010	EC24D1-010	2163	B1	-:H37
EC24D2-038-5	Dog	D2	038	EC24D2-038	101	B1	O93:H8
EC24H1-049-4	Human	H1	049	EC24H1-049	73	B2	O22:H1
EC24H1-053-4	Human	H1	053	EC24H1-053	58	B1	-:H37
EC23D1-044-1	Dog	D1	044	EC23D1-044	372	B2	O75:H31

Appendix 7: Metadata for the 217 *E. coli* genomes, isolated from humans and dogs including the Host ID, Household ID, Sample, MLST classification and Phylogroup classification are included.

Genome	Host	Host ID	Household	Sample	ST	Phylogroup	Serotype
EC23D1-044-2	Dog	D1	044	EC23D1-044	372	B2	O75:H31
EC23D1-044-3	Dog	D1	044	EC23D1-044	372	B2	O75:H31
EC23D1-044-4	Dog	D1	044	EC23D1-044	1079	B1	O6:H49
EC23D1-046-2	Dog	D1	046	EC23D1-046	942	B1	O163:H28
EC23D1-046-3	Dog	D1	046	EC23D1-046	131	B2	O25:H4
EC23D2-044-1	Dog	D2	044	EC23D2-044	10	A	O116:H20
EC23D2-044-2	Dog	D2	044	EC23D2-044	10	A	O116:H20
EC23D2-044-3	Dog	D2	044	EC23D2-044	10	A	O116:H20
EC23D2-044-4	Dog	D2	044	EC23D2-044	10	A	O116:H20
EC23H1-056-1	Human	H1	056	EC23H1-056	131	B2	O16:H5
EC23H1-056-2	Human	H1	056	EC23H1-056	963	D	O2:H18
EC23H1-056-3	Human	H1	056	EC23H1-056	131	B2	O16:H5
EC23H1-056-4	Human	H1	056	EC23H1-056	963	D	O2:H18

Appendix 8: Significantly associated genes and genetic variants associated with humans and dogs identified by both unitig Genome Wide Association Study (GWAS) and the corresponding pangenome cluster of orthologous genes (COG) ID

Gene or contig ID	Corresponding pangenome COG	hits	max p-value	Average allele frequency	Average beta	Approach
EC23D1-001-1_02244	group_957	6	7.62	0.328	-0.554	Unitig
EC23D1-001-1_02243|EC23D1-001-1_02244	group_1432 | group_957	3	9.05	0.117	-0.625	Unitig
flu_1	group_118	10	9.79	0.244	-0.562	Unitig
EC23D1-001-1_04881	group_2013	1	7.53	0.134	-0.599	Unitig
era_2	group_542	4	7.88	0.307	-0.424	Unitig
EC23D1-001-1_04800	group_2866	1	7.81	0.0922	-0.668	Unitig
EC23D1-001-1_02245	group_3507	1	7.29	0.0968	-0.645	Unitig
EC23D1-001-1_02243	group_1432	2	7.32	0.0967	-0.638	Unitig
EC23D1-001-1_04706	group_5667	2	7.57	0.477	0.576	Unitig
kilR	group_11760	1	7.74	0.235	-0.56	Unitig
EC23D1-001-1_04799|EC23D1-001-1_04800	group_6441 | group_2866	2	7.18	0.203	-0.47	Unitig
EC23D1-003-1_04759	group_461	1	7.53	0.24	0.495	Unitig
EC23D1-003-1_04836	group_2562	2	8.79	0.0645	-0.785	Unitig
EC23D1-003-1_01472	group_166	1	7.34	0.171	-0.583	Unitig
EC23D1-003-2_03458	group_9836	6	8.3	0.389	0.392	Unitig

Appendix 8: Significantly associated genes and genetic variants associated with humans and dogs identified by both unitig Genome Wide Association Study (GWAS) and the corresponding pangenome cluster of orthologous genes (COG) ID

Gene or contig ID	Corresponding pangenome COG	hits	max p-value	Average allele frequency	Average beta	Approach
traC_2	traC	5	7.89	0.357	0.384	Unitig
traR_2	traR_2	1	7.22	0.382	0.369	Unitig
EC23D1-003-2_03450	group_5575	1	8.46	0.359	0.414	Unitig
traD	traD	5	8.54	0.434	0.428	Unitig
EC23D1-003-2_03468	group_2084	5	8.56	0.33	0.411	Unitig
EC23D1-003-2_03462	group_5578	2	8.56	0.362	0.388	Unitig
ssb_2	ssb_3	1	7.28	0.424	0.428	Unitig
EC23D1-003-2_03451	group_5576	2	7.91	0.256	0.412	Unitig
ylpA	ylpA	2	7.78	0.415	0.381	Unitig
traV	group_2919	2	7.84	0.392	0.38	Unitig
yeeR	yeeR	1	7.16	0.258	0.57	Unitig
EC23D1-007-2_04580	group_501	1	7.19	0.0922	-0.908	Unitig
EC23D1-009-1_01060|klcA	group_2105 | group_2105	1	8.46	0.0553	0.861	Unitig
flu	group_118	5	9.23	0.127	-0.706	
EC23D1-012-1_05009	group_1695	2	8.04	0.0784	-0.626	Unitig
EC23D1-012-1_04878	group_3824	1	7.95	0.194	-0.628	Unitig
EC23D1-012-1_04906	group_4959	2	11.7	0.0737	-0.881	Unitig
cbtA_1	group_423	1	7.24	0.0323	-0.882	Unitig
EC23D1-012-2_02874	group_171	2	7.7	0.15	-0.586	Unitig
EC23D1-016-1_04326	group_2083	8	9.35	0.318	0.418	Unitig
EC23D1-016-1_04333	group_5578	3	8.08	0.286	0.409	Unitig

Appendix 8: Significantly associated genes and genetic variants associated with humans and dogs identified by both unitig Genome Wide Association Study (GWAS) and the corresponding pangenome cluster of orthologous genes (COG) ID

Gene or contig ID	Corresponding pangenome COG	hits	max p-value	Average allele frequency	Average beta	Approach
EC23D1-016-1_04329|traQ	group_4813| group_4813	2	8.67	0.256	0.436	Unitig & Pangenome
EC23D1-016-1_04335	group_2603	1	8.19	0.267	0.423	Unitig
EC23D1-016-1_04338	group_9836	1	8.27	0.249	0.425	Unitig
flu_2	group_120	1	7.69	0.166	-0.515	Unitig
era_4	era_2	1	7.24	0.0507	-0.931	Unitig
EC23H2-016-1_05231	group_4975	2	7.52	0.0138	0.99	Unitig
PID-1768-EC23D1-046-4_04070	group_2566	2	7.32	0.0276	-0.909	Unitig
PID-1768-EC23D1-046-4_04071	group_2568	5	7.32	0.0276	-0.909	Unitig
repB	repB	2	7.32	0.0276	-0.909	Unitig

Appendix 9: Significantly associated genes associated with humans and dogs identified by the pangenome cluster of orthologous genes (COG) presence and absence genome wide association study (GWAS) approach

Pangenome COG	Allele frequency	Likelihood ratio test p-value	Beta
acuR	0.0876	0.00000557	3.56
ampG_1	0.558	0.000000248	2.21
angR	0.553	0.00000093	2.04
araB_2	0.0783	0.000000445	3.97
bcsE	0.829	1.35E-11	-4.44
cesT	0.0645	0.00000499	-3.69
cotSA	0.207	0.000000423	-2.43
csbX_1	0.0783	0.000000445	3.97
csbX_2	0.0783	0.000000445	3.97
dhbE	0.558	0.000000248	2.21
ecpD	0.797	0.000000121	-2.56
ehaG	0.3	0.00000117	2.09
epsM_1	0.106	0.00000306	3.74
fyuA	0.553	0.00000093	2.04
gatA	0.871	0.000000835	-3.65
gatD	0.88	0.000000441	-3.98
gatY	0.88	0.000000441	-3.98
group_10008	0.138	0.00000533	-2.36
group_10015	0.207	0.000000423	-2.43
group_1031	0.0876	0.00000439	3.08
group_11238	0.106	0.00000306	3.74

Appendix 9: Significantly associated genes associated with humans and dogs identified by the pangenome cluster of orthologous genes (COG) presence and absence genome wide association study (GWAS) approach

Pangenome COG	Allele frequency	Likelihood ratio test p-value	Beta
group_1133	0.249	0.00000507	-1.61
group_11700	0.558	0.000000248	2.21
group_11702	0.558	0.000000248	2.21
group_1175	0.0691	0.0000052	-3.69
group_11844	0.12	0.0000005	-2.7
group_13090	0.207	0.000000423	-2.43
group_13096	0.0645	0.00000499	-3.69
group_1435	0.0922	0.000000149	-3.43
group_14823	0.203	0.00000172	2.28
group_15264	0.175	1.61E-09	4.24
group_1682	0.129	0.00000318	2.87
group_1711	0.0922	0.00000128	-3.04
group_17630	0.0783	0.000000445	3.97
group_17631	0.0783	0.000000445	3.97
group_18060	0.106	0.00000306	3.74
group_18065	0.106	0.00000306	3.74
group_1961	0.184	0.000000217	-2.45
group_2108	0.18	0.00000154	2.47
group_2292	0.207	0.00000341	2.05
group_2521	0.304	0.0000012	2.32
group_3029	0.0645	0.0000035	-3.74

Appendix 9: Significantly associated genes associated with humans and dogs identified by the pangenome cluster of orthologous genes (COG) presence and absence genome wide association study (GWAS) approach

Pangenome COG	Allele frequency	Likelihood ratio test p-value	Beta
group_3874	0.0645	0.0000035	-3.74
group_3895	0.0783	0.00000639	3.61
group_4075	0.157	3.58E-10	4.24
group_4813	0.217	0.00000364	1.92
group_5025	0.0876	0.00000563	3.55
group_5068	0.272	0.00000114	-2.1
group_586	0.396	8.55E-08	-1.87
group_6421	0.0691	0.00000036	-4
group_6700	0.138	0.00000533	-2.36
group_6706	0.0645	0.00000499	-3.69
group_6707	0.0645	0.00000499	-3.69
group_6731	0.249	0.00000337	2
group_6762	0.129	1.33E-08	4.02
group_7126	0.111	0.000000133	-3
group_8142	0.138	0.00000533	-2.36
group_8147	0.207	0.000000423	-2.43
group_8148	0.207	0.000000423	-2.43
group_8149	0.0645	0.00000499	-3.69
group_8150	0.0645	0.00000499	-3.69
group_8151	0.0645	0.00000499	-3.69
group_8153	0.0645	0.00000499	-3.69

Appendix 9: Significantly associated genes associated with humans and dogs identified by the pangenome cluster of orthologous genes (COG) presence and absence genome wide association study (GWAS) approach

Pangenome COG	Allele frequency	Likelihood ratio test p-value	Beta
group_8155	0.0645	0.00000499	-3.69
group_8157	0.0645	0.00000499	-3.69
group_8158	0.0645	0.00000499	-3.69
group_8159	0.0645	0.00000499	-3.69
group_8162	0.0645	0.00000499	-3.69
group_8163	0.0645	0.00000499	-3.69
group_8164	0.0645	0.00000499	-3.69
group_8165	0.0645	0.00000499	-3.69
group_8455	0.221	0.00000011	2.46
group_8868	0.111	0.000000133	-3
group_8870	0.0829	0.00000146	-2.92
group_9445	0.0553	0.00000371	-3.86
group_959	0.101	0.00000443	3.6
group_9682	0.189	0.00000249	-1.92
group_9886	0.235	0.000000516	-2.33
gtf1	0.0829	0.00000146	-2.92
htrL	0.226	1.38E-08	-2.67
hxpB_2	0.0783	0.000000445	3.97
irtA	0.558	0.000000248	2.21
legI	0.106	0.00000306	3.74
lpfB	0.189	0.00000249	-1.92

Appendix 9: Significantly associated genes associated with humans and dogs identified by the pangenome cluster of orthologous genes (COG) presence and absence genome wide association study (GWAS) approach

Pangenome COG	Allele frequency	Likelihood ratio test p-value	Beta
lpfC'	0.189	0.00000249	-1.92
lpfE	0.189	0.00000249	-1.92
lsrR_1	0.0783	0.000000445	3.97
malI_2	0.0783	0.000000445	3.97
mbtI	0.558	0.000000248	2.21
mpaA	0.737	6.11E-08	2.47
neuA	0.106	0.00000306	3.74
neuC	0.106	0.00000306	3.74
parE1	0.253	0.00000138	-2.13
pchR	0.558	0.000000248	2.21
pikAV	0.558	0.000000248	2.21
rbtD	0.0783	0.000000445	3.97
rfbC	0.0829	0.00000146	-2.92
rusA_3	0.217	0.000000168	2.4
spaP_2	0.0645	0.00000499	-3.69
spaR_3	0.0645	0.00000499	-3.69
spiA	0.0645	0.00000499	-3.69
srlR_1	0.88	0.000000441	-3.98
ssaV	0.0645	0.00000499	-3.69
traS	0.0968	0.000000131	3.97
upaG	0.488	9.41E-08	-1.72

Appendix 9: Significantly associated genes associated with humans and dogs identified by the pangenome cluster of orthologous genes (COG) presence and absence genome wide association study (GWAS) approach

Pangenome COG	Allele frequency	Likelihood ratio test p-value	Beta
vatD	0.267	0.00000183	-2.07
wbjC	0.101	0.000000133	-3
wfgD_2	0.0829	0.00000146	-2.92
wzxC_2	0.0829	0.00000146	-2.92
ybbA	0.138	0.00000533	-2.36
ymgF	0.618	0.00000387	-2.47
yscJ	0.0645	0.00000499	-3.69
yscN	0.0645	0.00000499	-3.69
yscU	0.0645	0.00000499	-3.69

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
EC23D1-001-1	
EC23D1-001-2	
EC23D1-003-1	
EC23D1-003-2	
EC23D1-006-1	
EC23D1-006-2	

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
EC23D1-007-1	
EC23D1-007-2	blaTEM_1A; catB3; dfrA1
EC23D1-009-1	
EC23D1-009-2	
EC23D1-010-1	
EC23D1-010-2	
EC23D1-011-1	
EC23D1-012-1	blaCMY_2
EC23D1-012-2	
EC23D1-013-1	
EC23D1-013-2	
EC23D1-016-1	blaTEM_1B; tet_B
EC23D1-016-2	blaTEM_1B; tet_B
EC23D1-023-1	
EC23D1-026-1	
EC23D1-026-2	
EC23D1-030-1	parC p.S57T; parE p.I355T
EC23D1-030-2	parC p.S57T; parE p.I355T
EC23D1-031-1	
EC23D1-031-2	
EC23D1-033-1	gyrA p.D87G
EC23D1-033-2	
EC23D1-038-1	parC p.S57T
EC23D1-038-2	parC p.S57T

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	**AMR genotype**
EC23D1-045-1	parC p.S57T; parE p.I355T
EC23D1-045-2	parC p.S57T; parE p.I355T
EC23D1-046-1	
EC23D1-047-1	
EC23D1-047-2	
EC23D1-048-1	aac_3__IId; blaTEM_1B; parE p.I529L
EC23D1-048-2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23D1-049-1	
EC23D1-049-2	
EC23D1-053-1	
EC23D1-053-2	
EC23D1-056-1	aac_3__IId; blaTEM_1B; parE p.I529L
EC23D1-056-2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23D2-038-1	parC p.S57T; parE p.I355T
EC23D2-038-2	
EC23D3-044-1	
EC23D3-044-2	
EC23H1-001-1	dfrA5; sul1
EC23H1-001-2	dfrA5; sul1
EC23H1-003-1	
EC23H1-003-2	
EC23H1-006-1	
EC23H1-006-2	
EC23H1-007-1	

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	**AMR genotype**
EC23H1-007-2	
EC23H1-009-1	
EC23H1-009-2	
EC23H1-010-1	parE p.I355T
EC23H1-010-2	
EC23H1-011-1	
EC23H1-011-2	
EC23H1-012-1	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; cmlA1; dfrA12; floR; sul2; sul3; tet_A
EC23H1-012-2	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA12; floR; sul2; sul3; tet_A
EC23H1-013-1	aph_3____Ib; aph_6__Id; sul2; tet_A
EC23H1-013-2	
EC23H1-016-1	erm_B; gyrA p.S83L
EC23H1-016-2	erm_B; gyrA p.S83L
EC23H1-023-1	aadA1; blaTEM_1B; dfrA1; sul1; tet_A
EC23H1-023-2	aadA1; blaTEM_1B; dfrA1; sul1; tet_A
EC23H1-026-1	blaTEM_1B; gyrA p.S83L; parE p.I529L
EC23H1-026-2	blaTEM_1B; gyrA p.S83L; parE p.I529L
EC23H1-030-1	
EC23H1-030-2	
EC23H1-031-1	
EC23H1-031-2	
EC23H1-033-1	parC p.S57T; parE p.I355T
EC23H1-033-2	parC p.S57T; parE p.I355T
EC23H1-038-1	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA7; qnrB19; sul1; sul2; tet_A

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
EC23H1-038-2	
EC23H1-044-1	aph_6__Id; blaCTX_M_15; blaTEM_1B; dfrA14; qnrS1; sul2; tet_A
EC23H1-044-2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23H1-045-1	parE p.I529L
EC23H1-045-2	parE p.I529L
EC23H1-046-1	qnrS1
EC23H1-046-2	qnrS1
EC23H1-047-2	parE p.I529L
EC23H1-048-1	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
EC23H1-048-2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23H1-049-1	aadA1; sul1
EC23H1-049-2	aadA1; sul1
EC23H1-053-1	aac_3__IId; aadA1; aadA2b; blaTEM_1B; cmlA1; dfrA12; qnrS1; sul3; tet_B
EC23H1-053-2	aac_3__IId; aadA1; aadA2b; blaTEM_1B; cmlA1; dfrA12; qnrS1; sul3; tet_B
EC23H2-003-1	
EC23H2-003-2	
EC23H2-006-1	
EC23H2-006-2	
EC23H2-010-1	
EC23H2-016-1	blaTEM_1B; tet_B
EC23H2-016-2	blaTEM_1B; tet_B
EC23H2-046-1	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA8; sul2; tet_B
EC23H2-046-2	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA8; sul2; tet_B

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	**AMR genotype**
EC23H2-048-1	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
EC23H2-048-2	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
EC23H3-045-1	
EC23H3-045-2	
PID-1768-EC23D1-001-3	
PID-1768-EC23D1-001-4	
PID-1768-EC23D1-003-3	
PID-1768-EC23D1-006-3	
PID-1768-EC23D1-006-4	
PID-1768-EC23D1-007-3	blaTEM_1A; catB3; dfrA1
PID-1768-EC23D1-007-4	blaTEM_1A; catB3; dfrA1
PID-1768-EC23D1-007-5	
PID-1768-EC23D1-009-3	parC p.S57T
PID-1768-EC23D1-009-4	parC p.S57T
PID-1768-EC23D1-012-3	
PID-1768-EC23D1-012-4	
PID-1768-EC23D1-013-3	
PID-1768-EC23D1-013-4	
PID-1768-EC23D1-013-5	
PID-1768-EC23D1-013-6	
PID-1768-EC23D1-016-3	blaTEM_1B; tet_B
PID-1768-EC23D1-016-4	blaTEM_1B; tet_B

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
PID-1768-EC23D1-026-3	
PID-1768-EC23D1-030-4	parC p.S57T; parE p.I355T
PID-1768-EC23D1-031-3	23S r.2032T>C
PID-1768-EC23D1-031-4	
PID-1768-EC23D1-033-3	
PID-1768-EC23D1-038-3	parC p.S57T
PID-1768-EC23D1-038-4	parC p.S57T
PID-1768-EC23D1-045-3	parC p.S57T; parE p.I355T
PID-1768-EC23D1-045-4	parC p.S57T; parE p.I355T
PID-1768-EC23D1-046-4	parE p.I529L
PID-1768-EC23D1-047-3	
PID-1768-EC23D1-047-4	
PID-1768-EC23D1-049-3	
PID-1768-EC23D1-053-3	
PID-1768-EC23D1-053-4	
PID-1768-EC23D1-056-3	gyrA p.D87N
PID-1768-EC23D1-056-4	gyrA p.D87N
PID-1768-EC23D2-038-4	parC p.S57T; parE p.I355T
PID-1768-EC23D3-044-3	
PID-1768-EC23D3-044-4	
PID-1768-EC23H1-001-3	dfrA5; sul1
PID-1768-EC23H1-001-4	dfrA5; sul1
PID-1768-EC23H1-003-3	
PID-1768-EC23H1-003-4	

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
PID-1768-EC23H1-006-3	
PID-1768-EC23H1-006-4	
PID-1768-EC23H1-007-3	
PID-1768-EC23H1-007-4	
PID-1768-EC23H1-009-3	
PID-1768-EC23H1-009-4	
PID-1768-EC23H1-010-3	parE p.I355T
PID-1768-EC23H1-010-4	
PID-1768-EC23H1-011-3	
PID-1768-EC23H1-011-4	
PID-1768-EC23H1-012-4	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; cmlA1; dfrA12; floR; sul3; tet_A
PID-1768-EC23H1-012-6	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; cmlA1; dfrA12; floR; sul2; sul3; tet_A
PID-1768-EC23H1-013-3	
PID-1768-EC23H1-013-4	aph_3____Ib; aph_6__Id; sul2; tet_A
PID-1768-EC23H1-013-5	
PID-1768-EC23H1-013-6	aph_3____Ib; aph_6__Id; sul2; tet_A
PID-1768-EC23H1-016-3	erm_B; gyrA p.S83L
PID-1768-EC23H1-016-4	erm_B; gyrA p.S83L
PID-1768-EC23H1-023-3	
PID-1768-EC23H1-026-3	blaTEM_1B; gyrA p.S83L; parE p.I529L
PID-1768-EC23H1-026-4	blaTEM_1B; gyrA p.S83L; parE p.I529L
PID-1768-EC23H1-026-5	blaTEM_1B; gyrA p.S83L; parE p.I529L
PID-1768-EC23H1-026-6	blaTEM_1B; gyrA p.S83L; parE p.I529L
PID-1768-EC23H1-03-5	

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
PID-1768-EC23H1-031-3	
PID-1768-EC23H1-031-4	
PID-1768-EC23H1-031-5	
PID-1768-EC23H1-033-3	dfrA5; tet_A; gyrA p.S83L
PID-1768-EC23H1-033-4	dfrA5; tet_A; gyrA p.S83L
PID-1768-EC23H1-038-3	
PID-1768-EC23H1-038-4	
PID-1768-EC23H1-044-3	
PID-1768-EC23H1-044-4	
PID-1768-EC23H1-044-5	aph_6__Id; blaCTX_M_15; blaTEM_1B; dfrA14; qnrS1; sul2; tet_A
PID-1768-EC23H1-045-3	parE p.I529L
PID-1768-EC23H1-045-4	parE p.I529L
PID-1768-EC23H1-046-4	qnrS1
PID-1768-EC23H1-046-5	qnrS1
PID-1768-EC23H1-047-3	parE p.I529L
PID-1768-EC23H1-047-4	parE p.I529L
PID-1768-EC23H1-049-3	aadA1; sul1
PID-1768-EC23H1-053-3	aac_3__IId; aadA1; aadA2b; blaTEM_1B; cmlA1; dfrA12; qnrS1; sul3; tet_B
PID-1768-EC23H2-003-3	
PID-1768-EC23H2-003-4	
PID-1768-EC23H2-006-3	blaTEM_1B; tet_B
PID-1768-EC23H2-006-4	
PID-1768-EC23H2-016-3	aadA1; blaSHV_48; sul1
PID-1768-EC23H2-016-4	blaTEM_1B; tet_B

Appendix 10: Combination of genes and point mutations conferring putative AMR from the 217 *E. coli* isolated from humans and dogs

Genome	AMR genotype
PID-1768-EC23H2-016-5	aadA1; blaSHV_48; sul1
PID-1768-EC23H2-048-3	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
PID-1768-EC23H2-048-4	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
PID-1836-EC24D1-003-4R	
PID-1836-EC24D1-010-3R	
PID-1836-EC24D2-038-5R	
PID-1836-EC24H1-049-4R	aadA1; sul1
PID-1836-EC24H1-053-4R	aac_3__IId; aadA1; aadA2b; blaTEM_1B; cmlA1; dfrA12; qnrS1; sul3; tet_B
PID-1856-EC23D1-044-1R	
PID-1856-EC23D1-044-2R	
PID-1856-EC23D1-044-3R	
PID-1856-EC23D1-044-4R	
PID-1856-EC23D1-046-2R	
PID-1856-EC23D1-046-3R	parE p.I529L
PID-1856-EC23D2-044-1R	aph_3____Ib; aph_6__Id
PID-1856-EC23D2-044-2R	aph_3____Ib; aph_6__Id
PID-1856-EC23D2-044-3R	aph_3____Ib; aph_6__Id
PID-1856-EC23D2-044-4R	aph_3____Ib; aph_6__Id
PID-1856-EC23H1-056-1R	aac_3__IId; blaTEM_1B; parE p.I529L
PID-1856-EC23H1-056-2R	
PID-1856-EC23H1-056-3R	aac_3__IId; blaTEM_1B; parE p.I529L
PID-1856-EC23H1-056-4R	23S r.2032G>C

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
EC23D1-001-1	D1	001	EC23D1-001	961	B2	
EC23D1-003-1	D1	003	EC23D1-003	1442	B2	
EC23D1-003-2	D1	003	EC23D1-003	978	B2	
EC23D1-006-1	D1	006	EC23D1-006	3576	B1	
EC23D1-007-1	D1	007	EC23D1-007	642	B1	
EC23D1-007-2	D1	007	EC23D1-007	novel6	B1	blaTEM_1A; catB3; dfrA1
EC23D1-009-1	D1	009	EC23D1-009	8573	D	
EC23D1-010-1	D1	010	EC23D1-010	2163	B1	
EC23D1-011-1	D1	011	EC23D1-011	101	B1	
EC23D1-012-1	D1	012	EC23D1-012	2538	D	blaCMY_2
EC23D1-012-2	D1	012	EC23D1-012	589	B2	
EC23D1-013-1	D1	013	EC23D1-013	399	A	
EC23D1-016-1	D1	016	EC23D1-016	12	B2	blaTEM_1B; tet_B
EC23D1-023-1	D1	023	EC23D1-023	6164	B2	

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
EC23D1-026-1	D1	026	EC23D1-026	101	B1	
EC23D1-030-1	D1	030	EC23D1-030	7345	cladeV	parC p.S57T; parE p.I355T
EC23D1-031-1	D1	031	EC23D1-031	10	A	
EC23D1-033-1	D1	033	EC23D1-033	12	B2	gyrA p.D87G
EC23D1-033-2	D1	033	EC23D1-033	12	B2	
EC23D1-038-1	D1	038	EC23D1-038	106	D	parC p.S57T
EC23D1-045-1	D1	045	EC23D1-045	7236	cladeV	parC p.S57T; parE p.I355T
EC23D1-046-1	D1	046	EC23D1-046	942	B1	
EC23D1-047-1	D1	047	EC23D1-047	101	B1	
EC23D1-048-1	D1	048	EC23D1-048	131	B2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23D1-049-1	D1	049	EC23D1-049	1800	B1	
EC23D1-053-1	D1	053	EC23D1-053	5415	B1	
EC23D1-056-1	D1	056	EC23D1-056	131	B2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23D2-038-1	D2	038	EC23D2-038	2559	cladeV	parC p.S57T; parE p.I355T

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
EC23D2-038-2	D2	038	EC23D2-038	101	B1	
EC23D3-044-1	D3	044	EC23D3-044	1079	B1	
EC23D3-044-2	D3	044	EC23D3-044	164	B1	
EC23H1-001-1	H1	001	EC23H1-001	8125	A	dfrA5; sul1
EC23H1-003-1	H1	003	EC23H1-003	69	D	
EC23H1-006-1	H1	006	EC23H1-006	141	B2	
EC23H1-007-1	H1	007	EC23H1-007	141	B2	
EC23H1-009-1	H1	009	EC23H1-009	38	D	
EC23H1-010-1	H1	010	EC23H1-010	59	F	parE p.I355T
EC23H1-010-2	H1	010	EC23H1-010	4429	A	
EC23H1-011-1	H1	011	EC23H1-011	906	B1	
EC23H1-011-2	H1	011	EC23H1-011	1079	B1	
EC23H1-012-1	H1	012	EC23H1-012	710	A	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; cmlA1; dfrA12; floR; sul2; sul3; tet_A

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
EC23H1-012-2	H1	012	EC23H1-012	710	A	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA12; floR; sul2; sul3; tet_A
EC23H1-013-1	H1	013	EC23H1-013	95	B2	aph_3____Ib; aph_6__Id; sul2; tet_A
EC23H1-013-2	H1	013	EC23H1-013	95	B2	
EC23H1-016-1	H1	016	EC23H1-016	9265	D	erm_B; gyrA p.S83L
EC23H1-023-1	H1	023	EC23H1-023	10	A	aadA1; blaTEM_1B; dfrA1; sul1; tet_A
EC23H1-026-1	H1	026	EC23H1-026	131	B2	blaTEM_1B; gyrA p.S83L; parE p.I529L
EC23H1-030-1	H1	030	EC23H1-030	372	B2	
EC23H1-030-2	H1	030	EC23H1-030	404	B2	
EC23H1-031-1	H1	031	EC23H1-031	5295	A	
EC23H1-033-1	H1	033	EC23H1-033	5345	cladeIII	parC p.S57T; parE p.I355T
EC23H1-038-1	H1	038	EC23H1-038	10	A	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA7; qnrB19; sul1; sul2; tet_A
EC23H1-038-2	H1	038	EC23H1-038	70	D	

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
EC23H1-044-1	H1	044	EC23H1-044	542	A	aph_6__Id; blaCTX_M_15; blaTEM_1B; dfrA14; qnrS1; sul2; tet_A
EC23H1-044-2	H1	044	EC23H1-044	131	B2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23H1-045-1	H1	045	EC23H1-045	131	B2	parE p.I529L
EC23H1-046-1	H1	046	EC23H1-046	34	A	qnrS1
EC23H1-047-2	H1	047	EC23H1-047	131	B2	parE p.I529L
EC23H1-048-1	H1	048	EC23H1-048	1193	B2	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
EC23H1-048-2	H1	048	EC23H1-048	131	B2	aac_3__IId; blaTEM_1B; parE p.I529L
EC23H1-049-1	H1	049	EC23H1-049	73	B2	aadA1; sul1
EC23H1-053-1	H1	053	EC23H1-053	58	B1	aac_3__IId; aadA1; aadA2b; blaTEM_1B; cmlA1; dfrA12; qnrS1; sul3; tet_B
EC23H2-003-1	H2	003	EC23H2-003	3759	B1	
EC23H2-006-1	H2	006	EC23H2-006	141	B2	
EC23H2-006-2	H2	006	EC23H2-006	635	A	

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
EC23H2-010-1	H2	010	EC23H2-010	362	D	
EC23H2-016-1	H2	016	EC23H2-016	12	B2	blaTEM_1B; tet_B
EC23H2-046-1	H2	046	EC23H2-046	69	D	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA8; sul2; tet_B
EC23H2-048-1	H2	048	EC23H2-048	1193	B2	aph_3____Ib; aph_6__Id; blaTEM_1B; dfrA17; mph_A; sul2; tet_B; gyrA p.S83L; gyrA p.D87N; parC p.S80I; parE p.L416F
EC23H3-045-1	H3	045	EC23H3-045	101	B1	
EC23H3-045-2	H3	045	EC23H3-045	1800	B1	
PID-1768-EC23D1-009-3	D1	009	EC23D1-009	720	D	parC p.S57T
PID-1768-EC23D1-012-3	D1	012	EC23D1-012	5254	B2	
PID-1768-EC23D1-013-3	D1	013	EC23D1-013	7075	A	
PID-1768-EC23D1-026-3	D1	026	EC23D1-026	349	D	
PID-1768-EC23D1-030-4	D1	030	EC23D1-030	7347	cladeV	parC p.S57T; parE p.I355T
PID-1768-EC23D1-031-3	D1	031	EC23D1-031	10	A	23S r.2032T>C

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
PID-1768-EC23D1-046-4	D1	046	EC23D1-046	131	B2	parE p.I529L
PID-1768-EC23D1-056-3	D1	056	EC23D1-056	372	B2	gyrA p.D87N
PID-1768-EC23H1-012-4	H1	012	EC23H1-012	710	A	aadA1; aadA2; aph_3____Ib; aph_6__Id; blaTEM_1B; cmlA1; dfrA12; floR; sul3; tet_A
PID-1768-EC23H1-023-3	H1	023	EC23H1-023	372	B2	
PID-1768-EC23H1-033-3	H1	033	EC23H1-033	95	B2	dfrA5; tet_A; gyrA p.S83L
PID-1768-EC23H1-044-3	H1	044	EC23H1-044	164	B1	
PID-1768-EC23H1-044-4	H1	044	EC23H1-044	2163	B1	
PID-1768-EC23H2-006-3	H2	006	EC23H2-006	141	B2	blaTEM_1B; tet_B
PID-1768-EC23H2-016-3	H2	016	EC23H2-016	73	B2	aadA1; blaSHV_48; sul1
PID-1856-EC23D1-044-1R	D1	044	EC23D1-044	372	B2	
PID-1856-EC23D1-044-4R	D1	044	EC23D1-044	1079	B1	

Appendix 11: De-replicated 92 *E. coli* isolated from humans and dogs

Genome	Host ID	Household	Sample	ST	Phylogroup	AMR genotype
PID-1856-EC23D2-044-1R	D2	044	EC23D2-044	10	A	aph_3____Ib; aph_6__Id
PID-1856-EC23H1-056-1R	H1	056	EC23H1-056	131	B2	aac_3__IId; blaTEM_1B; parE p.I529L
PID-1856-EC23H1-056-2R	H1	056	EC23H1-056	963	D	
PID-1856-EC23H1-056-4R	H1	056	EC23H1-056	963	D	23S r.2032G>C