THE ROLE OF INFLAMMATION IN THE FORMATION OF MULTI-DRUG RESISTANT LINEAGES OF *E. COLI*

— By —

CHRISTOPHER CONNOR

A thesis submitted to the University of Birmingham for the degree of:

DOCTOR OF PHILOSOPHY

July 2021

Institute of Microbiology and Infection
School of Biosciences
College of Medical and Dental Sciences
University of Birmingham

Acknowledgements

I would like to acknowledge the Wellcome Trust for funding this work, as well as my supervisors Prof. Alan McNally and Prof. Robin May for providing guidance and encouragement. I would also like to thank my collaborators: Dr. Braedon McDonald and his group who helped me perform the mouse colonisation experiments that form a significant portion of this thesis. Lastly, I would like to thank the members of the McNally group, T102, friends and family for providing support throughout my PhD.

Abstract

E. coli exists as a commensal in the gastrointestinal tract of humans and other animals but can also cause severe disease in humans. Pathogenic *E. coli* that infect sites outside of the intestine are called extra-intestinal pathogenic *E. coli* (ExPEC) and are the most common cause of urinary tract infections (UTI). ExPEC can also cause soft tissue infections, blood stream infections and meningitis all of which can be severe life threatening diseases. ExPEC infections are further complicated by antimicrobial resistance (AMR) with ExPEC infections increasingly presenting as multi-drug resistant (MDR) limiting the number of effective therapies. Curiously AMR in ExPEC is not uniform, instead specific clones are responsible for the majority of resistant infections. For example the spread of the AMR gene CTX-M has been attributed to a single ExPEC clone called ST131 which frequently causes UTI in humans. These MDR clones are adept at colonising healthy individuals with international travellers frequently importing MDR *E. coli* upon their return, resulting in onward transmission within households. Understanding why AMR is concentrated in specific clones and how these clones transmit between healthy individuals can help combat the spread of MDR pathogens. Previous pangenomic analysis has identified that genes involved in anaerobic metabolism exhibit increased variation in ST131 compared to other ExPEC lineages. Expanding on this previous analysis we identify a significant link between metabolism and carriage of AMR genes, specifically MDR lineages of *E. coli* display an increased abundance of genes associated with core energy metabolism and carbohydrate utilisation. Invading pathogenic bacteria are known to compete with resident gut commensals for limited nutri-

ents and so have evolved mechanisms to out-compete commensals. The host inflammatory response can inadvertently provide a novel set of compounds which enteric pathogens have adapted to exploit. We hypothesise that our observed metabolic signature in MDR ExPEC is evidence that they are also using compounds derived from the host inflammatory response. Testing this hypothesis we demonstrate that an MDR ST131 strain out-competes a commensal strain, isolated from a healthy human, in nutrient limited media under *in vitro* conditions. We further demonstrate that an MDR ST131 strain is able to out-compete a commensal strain when introduced into a germ-free mouse. The MDR ST131 is not only able to out-compete a commensal but also displace a resident commensal in a pre-colonised mouse. Displacement of the resident commensal occurs within 48 hours when mice are co-housed, with the ST131 strain becoming the dominant strain in all mice. Lastly, mice that are colonised by ST131 have elevated levels of pro-inflammatory cytokines in their caecum compared to commensally colonised mice. This thesis highlights the importance of metabolism in the evolution of MDR lineages of *E. coli* and contributes towards a greater understanding of how these pathogens spread.

Table of Contents

List of Figures

List of Tables

Acronyms

AMR Anti-Microbial Resistance

CDS Coding Sequence

CFU Colony Forming Units

COG Clusters of Orthologous Groups

DC Dendritic Cell

EAEC Enteroaggerative *E. coli*

EARS-Net European Antimicrobial Resistance Surveillance Network

EHEC Enterohaemorrhagic *E. coli*

EIEC Entero-Invasive *E. coli*

ELISA Enzyme-linked Immunosorbent Assay

EPEC Enteropathogenic *E. coli*

ESBL Extended Spectrum Beta-Lactamase

ETEC Enterotoxigenic *E. coli*

ExPEC Extra-intestinal Pathogenic *E. coli*

FQR Fluoroquinolone Resistant

GO Gene Ontology

HUS Haemolytic Uremic Syndrome

IGR Intergenic Region

ILC Innate Lymphoid Cell

iNOS Inducible Nitric Oxide Synthetase

InPEC Intestinal Pathogenic *E. coli*

LEE Locus of Enterocyte Effacement

LPS Lipopolysaccharide

LT Heat-Labile Toxin

MDR Multi-Drug Resistant

MLN Mesenteric Lymph Nodes

MLST Multi Locus Sequence Typing

MOI Multiplicity Of Infection

NK Natural Killer

NMEC Neonatal Meningitis *E. coli*

PAI Pathogenicity Associated Island

PAMP Pathogen Associated Molecular Pattern

PRR Pattern Recognition Receptor

RNS Reactive Nitrogen Species

ROD Region Of Difference

ROS Reactive Oxygen Species

SEPEC Sepsis associated *E. coli*

SNP Single Nucleotide Polymorphism

ST Sequence Type

ST Heat-Stable Toxin

Stx Shiga Toxin

T3SS Type-3 Secretion System

UPEC Urinary Pathogenic *E. coli*

VF Virulence Factor

XDR Extensively Drug Resistant

Chapter 1

Introduction

1.1 *Escherichia coli*

Escherichia coli is a gram-negative rod-shaped bacterium that was first described by Theodor Escherich in 1885. Since then it has become one of the most studied organisms on the planet being used as a model organism to understand aspects across biology from gene regulation to the process of evolution. *E. coli* is primarily found in the gastrointestinal tract of humans, birds, livestock and other animals but can also be found in the environment such as in water and sediment. *E. coli* can exist as a commensal in the gut of its host contributing to health through digestion of food substances, however it can also present itself as a pathogen causing severe disease in humans and other animals. The varied habitats and niches in which *E. coli* reside gives rise to a complex population structure.

1.2 Population Structure

1.2.1 Phylogroups & MLST

The environments in which *E. coli* has been identified are not homogenous, nor are its interactions with animals exhibiting both commensal and pathogenic behaviours. Therefore it is not surprising that the *E. coli* population is not homogenous and can be stratified into groups using a variety of phenotypic and genotypic techniques, with the most prevalent being serotyping, Multi Locus Sequence Typing (MLST) and phylotyping. Serotyping distinguishes *E. coli* based on surface molecules and whilst this technique is informative for how hosts respond to *E. coli* it is not informative of the evolutionary history of a strain. Serotyping is still used in specific contexts, particularly for discriminating pathogenic clones, however the technique has been superseded by phylotyping and MLST both of which leverage the power of DNA sequencing.

Phylotyping assigns isolates into one of 8 phylogroups that are defined using phylogenetic analysis of whole genome alignments. The groups are named A, B1, B2, C, D, E, F and G represent major clade divisions at the root of the *E. coli* phylogeny and are indicative of fundamental ancestral divergences in the population. The earliest division in the population is between phylogroups A, B1, C and E in one half and B2, F and D in the other (Figure 1.1). More recently additional groups G and H have been proposed [86], and as more sequence data becomes available it is possible that more phylogroups will be identified with some even proposing that certain phylogroups of *E. coli* are reclassified as sub-species or new species entirely [157, 37]. Phylotyping therefore identifies fundamental stratification of the *E. coli* population present at the root of the phylogeny. MLST on the other hand functions to separate individual clonal lineages within the phylogroups. This approach uses the combination of alleles present at specific loci in the genome to assign the bacteria to a Sequence Type (ST), with isolates in the same ST group considered to be more closely related to each

other than they are to isolates in other ST groups. Therefore ST groups are considered as clonal lineages which share a common ancestor. There are multiple MLST schemes with the most commonly used being the Warwick (Achtmann 7 gene) scheme, which largely dominated over the alternative Pasteur and Michigan schemes [28]. Each of these MLST schemes uses a different combination of genomic loci to determine ST groupings with only one loci shared between the 3 schemes. Generally, the loci used represent 'housekeeping' genes that are conserved across the *E. coli* population. The varying loci used by the schemes results in conflicting clustering of isolates. Specifically isolates assigned to ST131 by the Warwick MLST scheme are classified as ST43 under the Pasteur scheme, moreover the Pasteur scheme splits the Warwick ST131 lineage into two distinct groups. ST groups identified by MLST schemes can be hierarchically clustered to recapitulate the *E. coli* phylogeny. Comparing the phylogeny produced using MLST clustering and the one produced from whole genome alignment data determined that the Warwick scheme produced the closest match [130]. Whilst MLST methodology was devised before whole genome sequencing was readily achievable it is still of use, particularly in the identification and naming of pathogenic *E. coli* outbreak and pandemic clones. Ultimately, phylogroups represent ancestral divergences in the *E. coli* population that contain multiple individual clonal lineages which can be discriminated using MLST. Phylotyping is used to understand fundamental divergences in *E. coli* evolution and MLST is important for distinguishing between individual clonal lineages.

1.3 Commensal or Pathogen

E. coli is a normal resident of the human intestine, present in relatively low abundance, and contributes to host health through the digestion of food. However, *E. coli* is also capable of causing disease in humans and other animals. There are a variety of disease presentations in humans which are generally self-limiting and mild, however some can be severe and even fatal. Pathogenic *E. coli* are divided into groups called pathotypes or pathovars

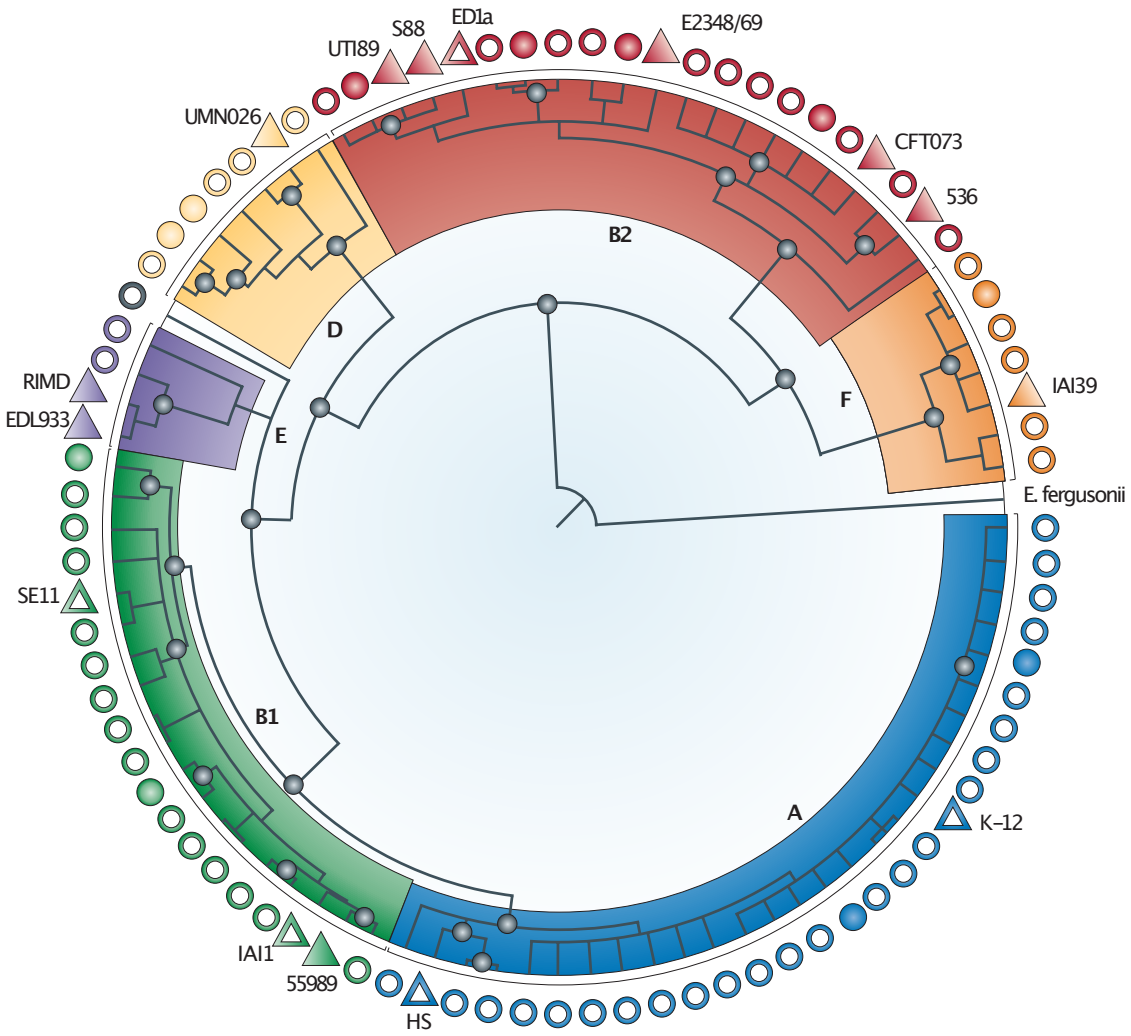

Fig. 1.1 Phylogenetic tree of 72 strains of *E. coli* rooted on *Escherichia fergusonii* - a closely related species of the same genus as *E. coli* that exhibits a broadly similar phenotype. *E. fergusonii* was included to help accurately resolve the structure of the *E. coli* phylogeny. Reference genomes are indicated by a triangle, pathogenic strains are denoted with closed symbols whilst commensals have open symbols. The major phylogroups are highlighted in coloured boxes. Figure reproduced from [145].

based on their disease presentation, pathogenesis and genetic factors. *E. coli* pathovars form two groups: those which cause disease in the intestinal tract (enteric pathogens) called Intestinal Pathogenic *E. coli* (InPEC) and those which cause disease outside of the intestine called Extra-intestinal Pathogenic *E. coli* (ExPEC). Comparative genomic analysis between commensal and pathogenic isolates has identified key genes involved in pathogenesis, these genes are referred to as Virulence Factors (VFs). Often these factors are clustered together in the genome in regions called Pathogenicity Associated Islands (PAIs) which tend to be located adjacent to tRNA genes [159, 134, 38]. *E. coli* pathogenesis has been extensively reviewed: [77, 37, 33].

1.3.1 Intestinal Pathogenic *E. coli* (InPEC)

This classification of *E. coli* includes those which cause enteric disease in humans with diarrhoea as the primary symptom, which can be either bloody or watery. Transmission between hosts occurs via the faecal oral route, in which consumption of food products contaminated with InPEC leads to intestinal colonisation and subsequent disease. Within InPEC there are numerous pathovars which can be distinguished based upon their clinical presentations, pathogenesis and the presence of specific VFs or PAIs.

Enteropathogenic *E. coli* (EPEC)

The first pathogenic *E. coli* were identified by Bray et al in 1945 and were originally named as diarrheagenic *E. coli* (DEC) [101], however it is now clear that these isolates belong to the Enteropathogenic *E. coli* (EPEC) pathotype. *E. coli* in this pathotype attach to intestinal epithelial cells using a variety of proteinaceous adhesion molecules called adhesins. Once closely anchored to the host epithelial cell membrane the bacterium uses secretion systems to inject effector proteins into the host cell. The transferred proteins induce abnormal rearrangements of the host cytoskeleton that produce a pedestal like protrusion which the anchored

E. coli sits atop. The manipulation of the host cytoskeleton results in loss of the epithelial microstructure (villi) and produces the characteristic attaching effacing (A/E) lesions which are the hallmark of EPEC infection. The close attachment and manipulation of the host epithelium stresses the intestinal epithelial cells which is detected by the host immune system, triggering an inflammatory response. Immune cells migrate to the site of infection in abundance and contribute to inflammation ultimately producing the diarrheagenic response. Mortality from these infections is low and generally only seen in infants, elderly or those with an impaired immune system. EPEC carry a PAI called the Locus of Enterocyte Effacement (LEE) which encodes numerous VFs required for pathogenesis, such as intimin for attachment to epithelial cells, Type-3 Secretion Systems (T3SSs) for transfer of effector proteins as well as the effector proteins themselves [92]. The LEE is not unique to EPEC and is found in another *E. coli* pathovar called Enterohaemorrhagic *E. coli* (EHEC).

Enterohaemorrhagic *E. coli* (EHEC)

EHEC also produce diarrhoea in the host which can be watery but is frequently bloody. This pathotype is extremely infectious with only a low number of ingested bacteria required to cause disease [146]. The intestinal tract of cows acts as a reservoir for this pathogen and outbreaks are generally associated with contaminated food products such as beef, but also fruits and vegetables which may have become contaminated by bovine faecal matter. The serotype O157 of EHEC is frequently identified as an outbreak strain associated with contaminated food products [https://www.cdc.gov/ecoli/outbreaks.html]. EHEC share the LEE PAI with EPEC but also produce a key toxin called Shiga Toxin (Stx). The Stx toxin causes damage to intestinal epithelial cells resulting in bloody diarrhoea and in severe cases perforation of the intestine. The toxin can also escape the intestinal lumen and act on distal organs with the microvasculature of the kidneys being particularly sensitive to Stx. Systemic action of the toxin results in Haemolytic Uremic Syndrome (HUS) characterised by a reduced

red blood cell count and impaired kidney fucntion. Children and the elderly are most at risk and while infections are typically self-limting, HUS can cause neurological damage, renal failure and death.

Enterotoxigenic *E. coli* (ETEC)

Enterotoxigenic *E. coli* (ETEC) use fimbrial colonisation factors to attach to the intestinal mucosa and produce toxins which interfere with host cellular processes. ETEC do not produce Stx but use a variety of other toxins which are divided into two groups called Heat-Labile Toxin (LT) and Heat-Stable Toxin (ST), based upon their thermostability. Both LT and ST interfere with signalling processes in the intestinal epithelial cells resulting in the disruption of ion homeostasis which ultimately leads to watery diarrhoea. Infants and travellers are most commonly affected but infections are generally self-limiting and non-fatal.

Enteroaggerative *E. coli* (EAEC)

Enteroaggerative *E. coli* (EAEC) are defined by their 'stacked brick' adhesion to epithelial cells which is evident in histopathological sections as clusters of self-aggregating bacteria, although it is unclear if this phenotype is unique to this pathovar [77]. EAEC, like ETEC, also produce toxins however EAEC do not produce LT or ST toxins nor do they produce Stx, instead they use a variety of autotransporter toxins to cause disease. Unlike the other InPEC pathovars there is no single VF or PAI that is present in all EAEC. Instead isolates rely on a combination of multiple different VFs, however the transcriptional regulator *aggR* is particularly common in EAEC and is associated with diarrhoeal disease [68]. The regulator controls the expression of multiple VFs found in EAEC.

Entero-Invasive *E. coli* (EIEC)

Unlike the other InPEC pathovars Entero-Invasive *E. coli* (EIEC) do not remain in the gut lumen attached to intestinal epithelial cells, instead EIEC invade the host epithelium. Initially they are contained within a vacuole inside the host cell however they use VFs to escape this compartment and become mobile through manipulation of the host cytoskeleton. Their mobility in the host cell allows them to spread into neighbouring epithelial cells. Despite their invasive capabilities EIEC rarely disseminate outside of the intestinal epithelium, although they are capable of evading the host immune response by inducing apoptosis in macrophages. The VFs responsible for pathogenesis in EIEC are encoded on a large virulence plasmid which can also be found in *Shigella* [131].

1.3.2 Extra-intestinal Pathogenic *E. coli* (ExPEC)

E. coli are not limited to causing disease in the intestine but are also capable of causing disease at extra-intestinal sites such as the urinary tract (UTI), kidneys, blood, soft tissues, lung and even brain. Initially *E. coli* were classified into pathovars such as: Urinary Pathogenic *E. coli* (UPEC), Sepsis associated *E. coli* (SEPEC) or Neonatal Meningitis *E. coli* (NMEC), however this classification was purely based upon disease presentation and to date no specific VFs have been identified that can discriminate between these pathovars. This lack of discriminative genetic factors led Russo et al. to propose the term Extra-Intestinal Pathogenic *E. coli* (ExPEC), which encompasses the variety of disease presentations outside of the intestine [129]. In contrast to InPEC, ExPEC do not cause disease when in the host intestine and can even be found in healthy individuals [104]. One meta-review estimated that at least 10% of the human population is asymptomatically colonised by an ExPEC strain [46]. However intestinal colonisation is thought to precede the development of disease, particularly UTIs, which are the most common ExPEC infection [34, 151, 36, 3, 99, 27, 104]. Near identical ExPEC strains have been isolated from both the urine and faecal samples of UTI patients

[99, 27, 104]. Moreover urine and faecal samples do not form discrete phylogenetic clusters [105], indicating a high degree of genetic similarity. Intestinal colonisation facilitates opportunistic transfer from the intestine to the urethra where ExPEC VFs allow adhesion to epithelial cells of the urinary tract (urothelium). Type 1 pili on *E. coli* are particularly important for this attachment [166, 100]. The pilus tip is encoded by *fimH* which binds to mannose epitopes on host cells. The interaction between fimH and mannose is dynamic: under low shear stress the attachment is weak permitting the bacteria to ascend the urinary tract however under high shear stress (such as during urine flow) the attachment strength increases dramatically [135]. Varying attachment strength depending on environmental forces allows *E. coli* to remain motile in the urinary tract without the risk of being expelled. Motility allows *E. coli* to ascend the urinary tract into the bladder where it can invade bladder epithelial cells to establish intracellular bacterial communities (IBC) causing cystitis [166]. Type 1 pili are again important for bacterial invasion into bladder epithelial cells, with fimH being necessary and sufficient for entry [91]. Bacteria can spread from the bladder to the kidneys by ascending the ureters where pilli again promote adhesion and invasion of epithelial cells causing pyelonephritis [97]. From the kidneys the bacteria can escape into the bloodstream causing bacteraemia and sepsis. Entry into the bloodstream also permits dissemination of bacteria to other organs and tissues.

ExPEC are therefore considered a facultative pathogen as they are capable of surviving in the host asymptomatically but under the right circumstances can cause disease. This highlights the clear distinction between InPEC and ExPEC with the former unable to exists as a commensal, producing disease in the host intestine while the latter can exist asymptomatically within the intestine and only causes disease outside of this compartment. This is reflected in the genomics of these different pathovars with InPEC possessing well defined and specific VFs such as the LEE of EHEC and the toxins of ETEC whilst ExPEC

lack a definitive VF. Some have speculated that pathogenicity in ExPEC is a by-product of increased colonisation efficiency [80].

1.3.3 Pathogen Genetics

There is no clear phylogenetic distinction between pathogenic and commensal *E. coli* with pathogenic clones observed in all phylogroups. Some pathovars do exhibit associations with specific phylogroups for example ExPEC clones can be found in multiple phylogroups but are predominantly found in B2, F and D whilst InPEC pathovars are predominantly found in phylogroups A and B1 [64]. The sporadic placement of pathogenic clones in the *E. coli* phylogeny indicates that pathovars have emerged repeatedly and independently in the population, and under the right conditions pathogenic clones can continue to arise. In fact there can be minimal differences between commensal and pathogenic strains, for example the commensal Nissle 1917 strain which is used probiotically contains remarkably few genetic differences to the virulent urinary isolate CFT073 [152]. Emergence of pathogenic *E. coli* clones is facilitated by the mobility of VFs and PAIs, which are frequently associated with mobile genetic elements. For example the VFs of EIEC are located on a plasmid, while the Stx toxin of EHEC is found on a phage. Analysis of both ETEC and EIEC has highlighted that both these pathovars have evolved repeatedly through the acquisition of mobile PAIs [156, 62]. Therefore genes which are circulating in the *E. coli* population can influence phenotype and shape the evolution of individual clones.

1.4 The Pangenome of *E. coli*

DNA sequencing technology allowed the complete genome of pathogenic and commensal clones of *E. coli* to be determined. Whilst the comparison between the commensal Nissle 1917 and pathogenic CFT073 strains revealed only minor differences this is not always the

case. The pathogenic EHEC O157 genome has numerous genetic differences compared to the commensal K12 MG1655 genome [118]. This comparison identified a multitude of genes involved in the pathogenesis of O157 including VFs and PAIs, however there were also numerous genes unique to O157 which had no clear role in virulence. Additionally, there were genes present in the commensal K12 that were absent in O157. Subsequently including an ExPEC genome in the analysis revealed that less than 40% of the genes identified were shared between all 3 isolates [158]. The lack of shared gene content between members of the same species gave rise to the concept of a pangenome, which is a representation of all the genes present in a selected population of genomes. The pangenome can be stratified into 2 compartments; the core genome which represents the genes present in all the members of that population; and the accessory genome which represents the genes which are not shared. Genes circulating in the accessory genome can be hugely influential on *E. coli* phenotype as it allows the bacteria to take a 'pick and mix' approach to evolution. This is particularly evident in new pathogenic clones of *E. coli* which have combined VFs from multiple pathovars. In 2011, an *E. coli* clone caused a severe outbreak in Germany which infected 3,816 people of which 845 developed HUS and 36 died. The clone had an O104:H4 serotype and combined pathogenic mechanisms from EHEC, EAEC and ExPEC [52, 124]. Specifically the strain produced the Stx toxin of EHEC (but lacked the LEE), it adhered to intestinal epithelial cells in a manner similar to EAEC and possessed multiple VFs typically found in ExPEC isolates such as iron acquisition systems. Between 2005 and 2014 another pathogenic *E. coli* clone emerged in France that had combined pathogenic traits from EHEC and ExPEC. The clone belonged to phylogroup A, which is an uncommon phylogroup for both ExPEC and EHEC, and was serotype O80 [140]. The clone was capable of causing both HUS and blood stream infections, traits which are typically reserved to EHEC and ExPEC respectively. The genome of this clone contained multiple EHEC VFs as well as a plasmid containing ExPEC VFs. Both of these pathogenic clones had combined VFs circulating in the *E. coli* accessory

genome to produce new successful phenotypes. The acquisition of genes circulating in the accessory pangenome, such as VFs, can therefore influence the evolution of *E. coli*.

1.5 Antibiotics

Antibiotics can be broadly divided into 2 categories depending on whether they directly kill bacteria (bactericidal), or inhibit bacterial growth (bacteriostatic). Antibiotics are further divided into classes depending upon their chemical structure and spectrum of activity. Classes of antibiotics include, but are not limited to, penicillins, fluoroquinolones, cephalosporins, carbapenems and sulphonamides. The variety of chemical structures gives rise to a variety of mechanisms of action against bacteria. For example fluoroquinolone antibiotics, such as ciprofloxacin, bind to DNA topoisomerase or DNA gyrase enzymes inhibiting their action. This prevents the bacterium from performing the structural alterations to its genome necessary for replication and transcription, ultimately inhibiting bacterial growth. Sulphonamides, such as sulfamethoxazole, interfere with the folate synthesis pathway by inhibiting the necessary metabolic enzymes, this starves the bacterium of this essential metabolite thereby preventing growth. Penicillins, cephalosporins and carbapenems all share a β-lactam ring in their chemical structure, this ring is essential for the mechanism of action as it acts as a substrate for cell wall synthesis enzymes thereby inhibiting them. Consequently, bacteria die when they are unable to synthesise or maintain their cell wall. The distinction between penicillins, cephalosporins and carbapenems is their original source. Specifically penicillins are derived from the penicillin molecule which was first identified from a fungus called *Penicillium*. The chemical structure of the compound has been altered over time giving rise to the penicillin family. Likewise both cephalosporins and carbapenems were first identified in microbes and were subsequently chemically altered to generate new families of antibiotics. Other classes of antibiotics also function to inhibit crucial bacterial processes such as protein synthesis and membrane integrity. Antibiotics have varying spectra of activity with some

being effective against a limited selection of bacteria whilst others, such as beta-lactams and fluoroquinolones, are effective against a broad spectrum of bacteria. Beta-lactam and fluoroquinolone antibiotics are some of the most widely used antibiotics against bacterial infections, including those caused by *E. coli*. Unfortunately bacteria are evolving mechanisms to resist the effects of antibiotics, termed Anti-Microbial Resistance (AMR).

1.6 Antibiotic Resistance

Bacteria are considered resistant when they are capable of surviving in the presence of an antibiotic compound. Resistance can be divided into intrinsic and acquired resistance, with the latter being a significant concern for modern medicine. Intrinsic resistance occurs when a bacteria is naturally resistant to an antibiotic, for example gram-negative bacteria are innately resistant to vancomycin as their outer membrane prevents the compound from reaching the cell wall. Acquired resistance necessitates the acquisition of genetic determinants that permit a previously susceptible bacteria to survive in the presence of an antibiotic. These genetic determinants can be acquired by mutation or through horizontal gene transfer. The acquired genetic determinants confer resistance by altering: antibiotic uptake / efflux, target sites, metabolic pathways or permitting degradation of the antibiotic.

1.6.1 Antibiotic Resistance Mechanisms

Antibiotic Uptake and Efflux

Antibiotics must be able to reach their target site to have any effect on bacteria, for example vancomycin is not effective against gram-negative bacteria as it is unable to pass through the outer membrane. Common targets for antibiotics are the cell wall, metabolic enzymes, ribosomes and DNA topoisomerases most of which are in the cytosol, therefore antibiotics must be able to penetrate bacterial membranes and cell wall to reach their target. Bacte-

ria use membrane bound proteins called porins which form channels in the cell membrane to absorb compounds from their environment, often antibiotics also enter bacteria through these pores. Alterations in porin structure or concentration can influence the permeability of the cell membrane and hence reduce antibiotic absorption conferring resistance [153]. For example mutations in the *ompF* gene, encoding a porin, in *E. coli* have been demonstrated to result in decreased absorbance of β-lactam antibiotics [138]. Bacteria are also capable of controlling the movement of compounds from their cytosol to the environment through efflux pumps. Increased removal of antibiotics from the cell through alterations to efflux pumps also functions to reduce antibiotic concentration within the cell conferring resistance. Generally altered membrane permeability through changes to porins only confers low levels of resistance but this can be sufficient to allow the bacteria to acquire more significant resistance mechanisms such as target site alteration.

Target Site Alterations

Antibiotics mediate their effect through the inhibition of essential bacterial cell processes typically by occupying or binding to a biologically active site. For example fluoroquinolone antibiotics inhibit the action of DNA topoisomerases preventing the bacterium from replicating its genome. Mutations in DNA topoisomerase genes that reduce fluoroquinolone binding allow these enzymes to function uninhibited by the presence of fluoroquinolone antibiotics. Resistance to sulfamethoxazole and trimethoprim can also occur via target site alterations. These antibiotics inhibit the enzymes dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR) both of which are essential for folate synthesis in the bacteria. Mutations in these enzymes reduce the affinity for antibiotics allowing these enzymes to continue folate synthesis. Resistance to these antibiotics can also occur via target site bypass.

Target Bypass

Bacteria can acquire resistance by bypassing the target of the antibiotic preventing it from inhibiting essential bacterial processes. While resistance to sulfamethoxazole and trimethoprim can be acquired via mutations bacteria can also acquire alternative alleles for DHPS and DHFR via horizontal gene transfer. These horizontally acquired alleles are insensitive to sulfamethoxazole / trimethoprim allowing bacterial metabolism to bypass the antibiotic target. Resistance to β-lactams can also occur via target site bypass with certain bacteria producing cell wall synthesis enzymes that can function in the presence of β-lactams. Resistance to β-lactams is more frequently mediated by antibiotic modification.

Antibiotic Modification

Bacteria can evolve resistance by acquiring enzymes which directly modify the antibiotic rendering it ineffective. Typically the bacteria produces an enzyme that adds a chemical group onto the antibiotic such as adding an acetyl or phosphate group. The modification results in reduced binding between the antibiotic and its target. For example resistance to aminoglycoside compounds is mediated by enzymes that add either a acetyl-, adenyl- or phosphate group onto the aminoglycoside. Bacteria can also modify antibiotics to remove their functional group, this is frequently the mechanism of resistance for β-lactam antibiotics. This group of antibiotics are named after the β-lactam ring they contain which is responsible for irreversible binding to cell wall synthesis enzymes. β-lactamase enzymes hydrolyse the β-lactam ring rendering the antibiotic harmless. β-lactamases have not evolved recently but instead have existed for billions of years however the use of β-lactam antibiotics clinically has certainly contributed to a recent diversification of this group of enzymes [58].

1.6.2 β-lactams, ESBLs and Multi-Drug Resistance

Shortly after the introduction of β-lactam antibiotics into clinical practice reports of resistance in pathogenic bacteria began to emerge. Further investigation led to the identification of β-lactamases which were destroying the antibiotic. As rates of resistance increased pharmaceutical companies searched for new compounds that were not destroyed by these enzymes giving rise to families of antibiotics named after their progenitor (penicillins, cephalosporins and carbapenems). Early β-lactamases only conferred resistance to a limited number of β-lactam antibiotics however as the diversity of β-lactam antibiotics in use increased β-lactamases evolved to have a broader spectrum of activity. β-lactamases which are able to destroy pencillins, 1st, 2nd and 3rd generation cephalosporins are called Extended Spectrum Beta-Lactamases (ESBLs), fortunately these enzymes are generally unable to degrade carbapenems and are inhibited by β-lactamase inhibitors [115]. ESBLs arose through mutations which affected the active site of the enzyme allowing it to degrade β-lactams with a variety of chemical side groups. Carbapenem antibiotics are considered the last effective β-lactam antibiotic against ESBL producing bacteria however carbapenemases are now emerging capable of degrading these antibiotics. Carbapenemases have arisen from ESBLs through further mutations which broaden their spectrum of activity [39]. The widespread use of β-lactam antibiotics and their reformulations has led to a huge diversity in β-lactamase enzymes. As of March 2021 there are 1,148 β-lactamase entries in the AMR database Resfinder [17]. Worryingly ESBLs can frequently be found on conjugative plasmids facilitating their transfer between bacteria. Mobile genetic elements, such as phage and insertion sequences that facilitate the movement of VFs, can also mobilise resistance genes. Mobile genetic elements are particularly important for transporting resistance genes from the chromosome onto conjugative plasmids. Ultimately this has resulted in plasmids which carry genes conferring resistance to a diverse set of antibiotic compounds. This convergence of multiple resistance mechanisms onto mobile plasmids has promoted the rise of bacterial pathogens that are re-

sistant to multiple compounds, these bacteria are referred to as Multi-Drug Resistant (MDR) [89]. This cumulative acquisition of resistance mechanisms threatens to produce pathogens for which there is no effective treatment. There are already reports of Extensively Drug Resistant (XDR) pathogens for which only one or two classes of antibiotic remain effective [89]. Resistance to antibiotics is not limited to certain countries or healthcare systems but is a global problem with rates of resistance increasing around the world.

1.7 Antimicrobial Resistance in ExPEC

Rates of resistance in *E. coli* have increased significantly since *circa* 2000 [127, 34, 25], particularly in ExPEC populations that cause UTI and blood stream infections. In 2015 antibiotic resistant *E. coli* were calculated to have the greatest impact on mortality and health compared to other major antibiotic resistant pathogens: being responsible for between 7787 and 10607 deaths across Europe, with rates varying by country [25]. As of 2019 antibiotic resistant *E. coli* were the most frequently reported invasive isolate causing 44.2% of blood stream infections in Europe. Of those isolates 57.1% were resistant to at least one antimicrobial being surveyed by the European Antimicrobial Resistance Surveillance Network (EARS-Net) [47]. The increase in resistance in *E. coli* is not restricted to Europe but has been observed globally [116, 10, 127]. ExPEC infections are typically treated with antibiotics from the fluoroquinolone or β-lactam classes however resistance to both has increased in recent decades. Resistance to β-lacatam antibiotics in ExPEC appears to have been conferred by aquisition of the ESBL CTX-M, whose prevalence has increased markedly in recent decades [127, 23].

Named after it's 'preference' for cefotaxime over ceftazidime (cephalosporins), CTX-M originated in the *Kluyvera* genus which is an environmental species but can occasionally cause infections in humans [126, 107]. The insertion sequence ISEcp1 mobilised the CTX-M gene moving it from the chromosome onto a plasmid [121], allowing it to subse-

quently transfer into *E. coli*. The presence of CTX-M confers resistance to third generation cephalosporins but not carbapenems. There is considerable diversity of CTX-M genes with alleles originating in different *Kluyvera* species and presenting with altered spectra of activity [24, 107, 126]. The alleles of CTX-M also display geographic variation with CTX-M-14 dominant in South East Asia, CTX-M-2 is more common in South America [16], whilst CTX-M-15 can be found worldwide [103, 30]. CTX-M was likely spread globally through international travel as CTX-M is the most common ESBL in strains colonising travellers to AMR endemic regions [144, 7]. The rapid emergence of CTX-M in clinical *E. coli* was strongly associated with the emergence of a single ExPEC clone: ST131 [103, 30].

1.7.1 *E. coli* ST131

In 2006 and 2007 two studies reported that CTX-M-15 was carried almost exclusively by one clonal lineage of *E. coli* identified under the Warwick MLST scheme as ST131 [23, 103]. Both studies observed that their collection of CTX-M positive isolates displayed an abundance of ST131 with a mixture of other STs forming the minority. Moreover both groups used a collection of isolates from around the globe, indicating that this clone was globally dispersed and probably responsible for the spread of CTX-M-15. Since the reported association between ST131 and CTX-M carriage others have investigated the prevalence of ST131 in wider collections of resistant clinical isolates. This led to the identification that not only was ST131 associated with carriage of CTX-M but was also overrepresented in collections of Fluoroquinolone Resistant (FQR) isolates [73, 151]. The emergence of ST131 in some surveillance datasets was coincident with a marked increase in resistance, to both β-lactam and fluoroquinolone antibiotics [73]. Subsequently, ST131 has repeatedly been identified as a major constituent in multiple collections of resistant clinical isolates from around the globe [116, 10, 3, 32]. It is therefore clear that ST131 is a globally disseminated lineage of *E. coli* that is strongly associated with antibiotic resistance, this has led some to

propose that the increase in ST131 is responsible for the increase in AMR in the ExPEC population [36, 73]. The evolution of ST131 has been the focus of numerous studies to understand why this lineage is a significant contributor to AMR.

1.8 Evolution of ST131

1.8.1 Nested Structure of H30, H30R and H30Rx

Initial investigations of the ST131 population revealed a significant association between the *fimH*30 allele (which encodes a type 1 fimbrial adhesin protein) and FQR [73]. The H30 allele was observed to be almost exclusively present in the FQR isolates, moreover the appearance of the H30 allele in the dataset was coincident with an increase in FQR. Later studies identified a number of H30 isolates that lacked FQR, which led the authors to postulate that the H30 allele was acquired before FQR [123]. This hypothesis was supported with phylogenetic analysis that clustered the FQR isolates as a single clade suggesting a single aquisition of FQR by a H30 ancestor [123]. The authors therefore proposed the H30R lineage defined by FQR that existed within the larger H30 clade [123]. Phylogenetic analysis also identified that CTX-M carrying isolates formed a single cluster within the H30R clone [123], the authors named this population H30Rx for its extended resistance profile. This identified a nested structure of the ST131 population where H30Rx was a single clade within a larger H30R clade which in turn was part of a larger H30 clade (Figure 1.2).

1.8.2 The Clades of ST131

Phylogenetic analysis of whole genome sequence data revealed that the ST131 lineage could be stratified into 3 clades named A, B and C [119]. Clade A has been estimated to have diverged from clades B/C first, *circa* 1874, with clade B and C diverging later with estimates ranging from 1960 to 1980 [143, 74, 11]. Clade C is further subdivided into C1 and C2

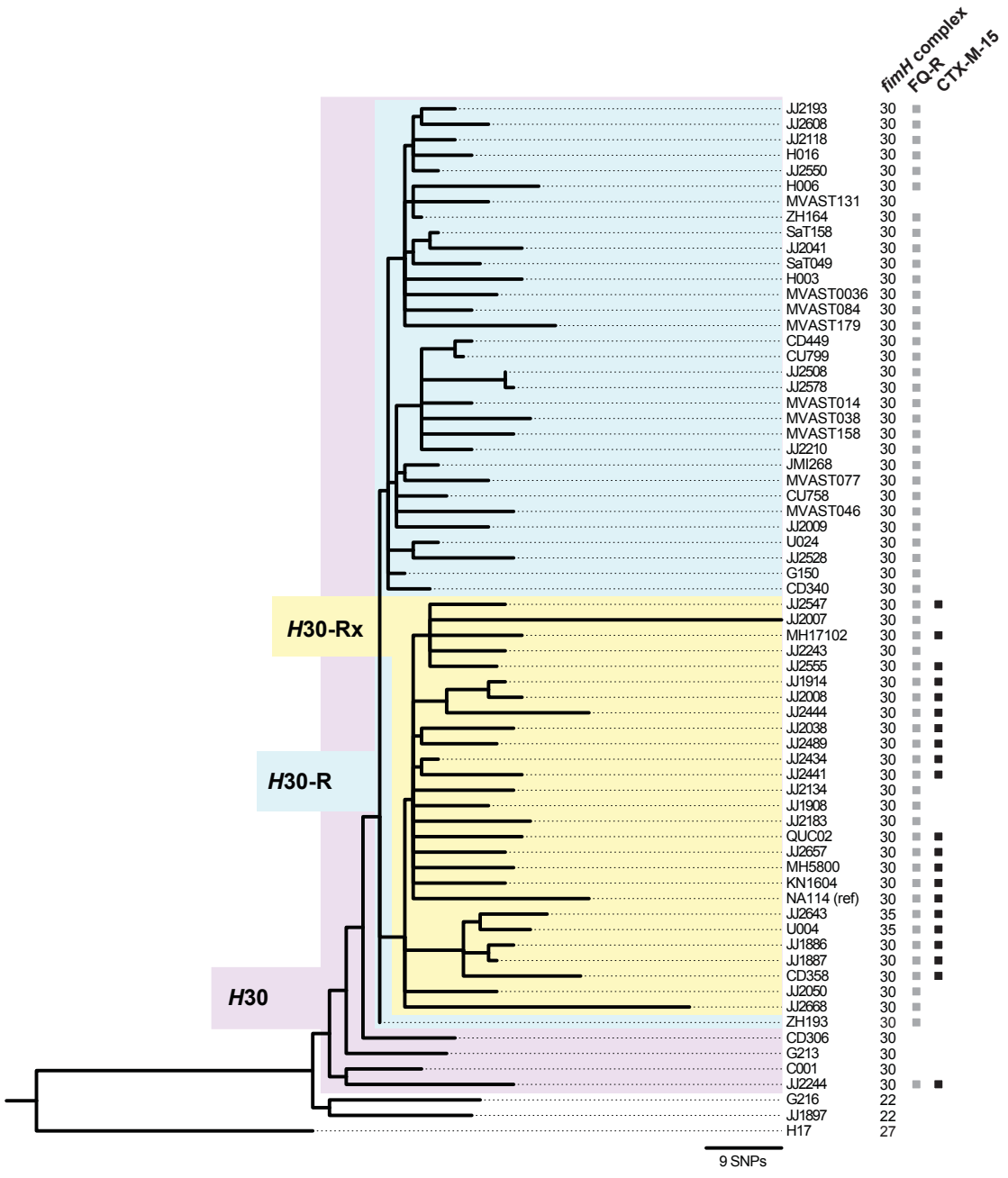

Fig. 1.2 Nested structure of H30, H30R and H30Rx of ST131. Figure reproduced from [123].

which are estimated to have diverged around 1987 [11]. The majority of isolates present in C1 displayed resistance to fluoroquinolones but lack CTX-M, corresponding to the previously identified H30R clone. The isolates in C2 display fluoroquinolone resistance and possess CTX-M corresponding to the H30Rx clone [119, 143]. The C2 clade also possesses more resistance genes in total compared to the other clades [11]. Intermediate clades have also been proposed due to the discovery of strains that possess intermediate genomic characteristics between clades B and C, these clades have been named as B0 and C0 [11]. ST131 therefore has a stratified population with 3 major clades A, B and C, the latter contains the majority of AMR and can be subdivided into C1/H30R and C2 /H30Rx (Figure 1.3).

1.8.3 Clonal Expanstion or Complex Evolution

The nested structure of increasing resistance within the H30 lineage implies a sequential series of evolutionary events, each of which led to the expansion of a higher fitness clone. Specifically a H30 clone emerged and acquired FQR promoting its expansion. One of the descendants then acquired CTX-M-15 which again promoted clonal expansion. Giving rise to the nested structure observed today. This hypothesis is supported by the observation that FQR across the ST131 lineage is mediated by the same set of Single Nucleotide Polymorphisms (SNPs) in *gyrA* and *parC* [73]. This combination of SNPs is restricted to ST131 and only infrequently occurs in the wider *E. coli* population [73]. In addition to the conservation of FQR SNPs across ST131, the sequence of CTX-M also displays remarkably little variation implying that all the alleles present in the lineage descended from a single copy [123]. Morevoer CTX-M is strongly associated with an IncFII plasmid in ST131 [143], suggesting a single acquisition of a CTX-M bearing plasmid. However, previous reports contradict this finding having observed no significant association between CTX-M and a common plasmid backbone, nor a consistent gene neighbourhood [119], leading to a conclusion that there were multiple independent aquisitions of CTX-M. Further parsimony analysis determined

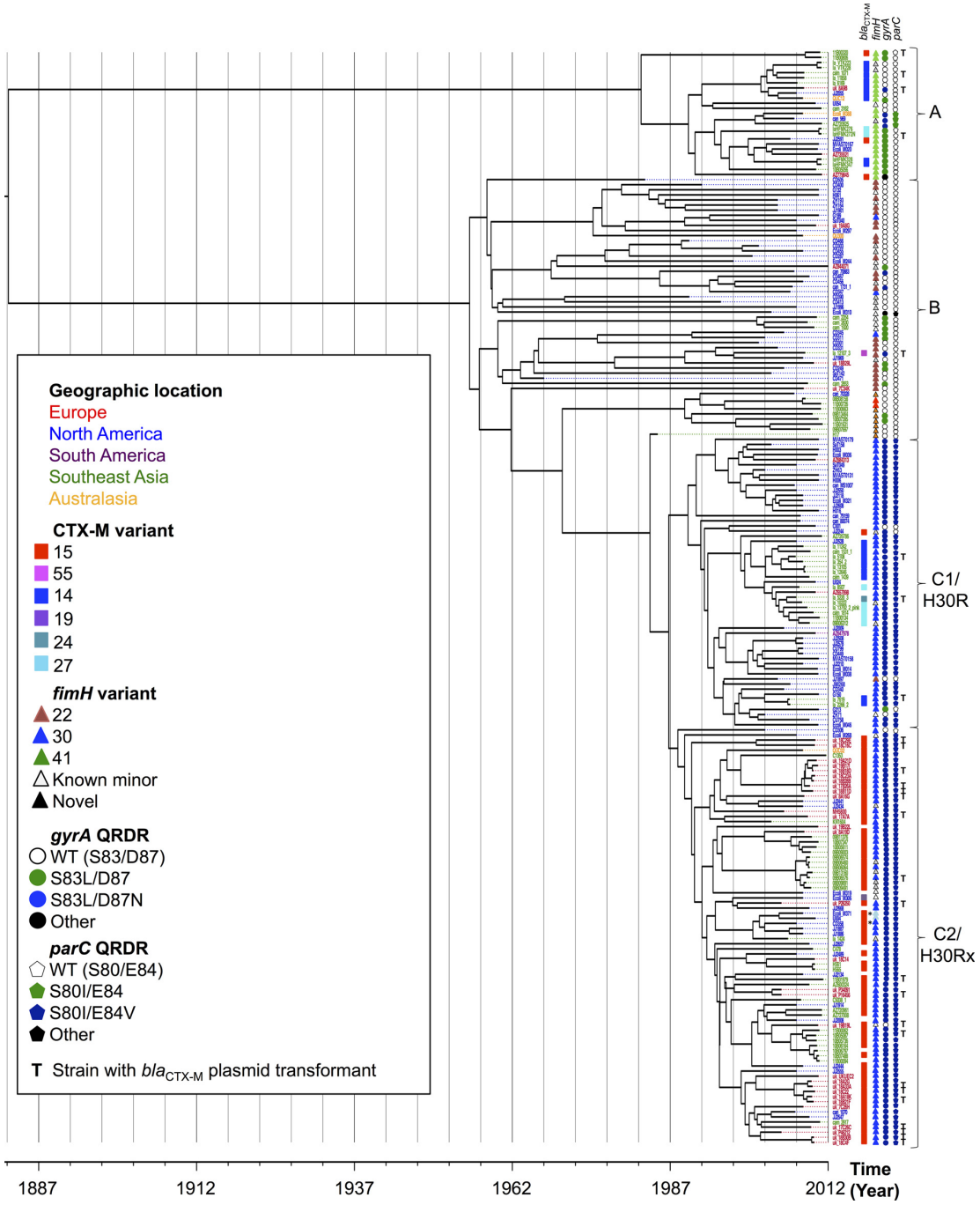

Fig. 1.3 Clades of ST131. The fimH30 allele (blue triangle) is concentrated in clade C as are the SNPs conferring FQR (blue circle and pentagon). CTX-M-15 (red square) is concentrated in clade C2. Figure reproduced from [143].

that there were 28 independent gain or loss events of CTX-M in ST131 [74]; inconsistent with a single acquisition event with subsequent clonal expansion. Lastly, the 3 clades of ST131 lack any discernible temporal or geographic association which would be expected from sequential clonal expansions [143, 11]. Ultimately the sequence of evolutionary events giving rise to ST131 are still debated.

1.8.4 What has made ST131 successful?

Determining how ST131 has acquired antibiotic resistance and spread globally is of great importance for understanding the evolution and spread of AMR. Initially the link between VFs and AMR was investigated due to the association between the *fimH30* allele and FQR in ST131 [9]. *FimH* encodes the tip of the type 1 pilus which is essential for ExPEC to adhere to and invade epithelial cells in the urinary tract [166, 100]. Genomic analysis indicated that ST131 acquired the *fimH30* allele along with PAIs at the *pheV* and *leuX* loci (tRNA genes) before FQR [11]. Leading the authors to propose that these acquired VFs were responsible for the success of ST131. However, the acquired VFs are not unique to ST131 and are well characterised in numerous ExPEC lineages that have not formed MDR clones. While principal component analysis indicated that ST131 possesses a distinctive set of VFs [70], these VFs are not unique to ST131 but found in other ExPEC lineages. ST131 does not possess any VFs or PAIs that distinguish this lineage from other ExPEC lineages, nor is there any difference in total number of VFs in ST131 compared to other major ExPEC lineages [32, 106]. Most importantly mouse models of ExPEC infections have failed to consistently demonstrate that ST131 is more virulent than other ExPEC [72]. While VFs are likely to be important for the success of ST131, the lack of a unique VF implies that they are not solely responsible for the success of this lineage. Other factors contribute to the success of this lineage, for example there is evidence that ST131 has adapted to offset the fitness costs imposed by AMR genes. Analysis of promoter regions in the ST131 lineage revealed a significant asso-

ciation between CTX-M alleles and specific promoter regions [96]. Comparative genomics and population pangenome analysis has also highlighted differences in metabolic gene content between ST131 and other ExPEC clones. For example early comparisons between an ST131 genome and a number of other *E. coli* genomes identified a number of genomic islands unique to ST131 termed Region Of Difference (ROD) [149], one of which contains a cluster of sugar metabolism genes. Analysis of the ST131 pangenome identified 754 loci that were significantly associated with ST131 compared to other ExPEC lineages, of these 292 were metabolic genes [96]. Further pangenome analysis employed a stringent identity threshold for clustering gene sequences, specifically if the protein sequence of two genes diverged by more than 5% the genes would be separated in the pangenome. This approach can discriminate between alleles of the same gene thereby identifying genes exhibiting variation in the ST131 lineage. This technique identified that genes associated with anaerobic metabolism were exhibiting increased variation in ST131 [95]. To determine whether this was simply a virulence trait exhibited by all ExPEC clones the pangenome of ST131 was compared to the pangenome of the other dominant ExPEC lineages ST73 and ST95, to account for differences in sample size the pangenome of ST131 was repeatedly subsampled and compared to ST73 and ST95. The results demonstrated that anaerobic metabolic gene diversity was significantly enriched in ST131, eliminating the possibility that this was a common ExPEC pathogenic trait. The identified alleles of the anaerobic metabolic genes were restricted to clades B and C of ST131, which is where AMR genes are concentrated [95]. This suggests that metabolism, particularly anaerobic metabolism, is important for the evolution of ST131.

1.8.5 Variation in anaerobic metabolism is shared amongst MDR ExPEC

Surveillance of MDR *E. coli* has identified that ST131 is not the sole MDR lineage, there are a limited number of other ExPEC lineages associated with a high carriage of AMR genes

such as ST167, ST410 and ST617 [172, 50, 169, 20]. These three lineages are of particular concern as they are involved in the global spread of carbapenemases, which confer resistance to carbapenem antibiotics that are considered the last line of defence against MDR ExPEC. Importantly these lineages also display signs that selective pressure is acting on anaerobic metabolism. ST167 and ST617 possess unique alleles of anaerobic dehydrogenase genes which are absent in the closely related commensal lineage: ST10 [172], while the ST410 lineage possesses a unique set of SNPs located in anaerobic metabolic genes [50]. Therefore the variation in anaerobic metabolism observed in ST131 appears to be shared with other MDR ExPEC lineages, suggesting that it is important for the evolution of successful AMR clones. Anaerobic metabolism is known to be crucial for the growth of *E. coli* in the inflamed gut [162, 141].

1.9 Gastrointestinal Inflammation & Anaerobic Metabolism

Bacteria liberate energy from a variety of compounds through respiration. A series of catabolic reactions drives the movement of electrons which powers the production of ATP. At the end of the electron transport chain is a terminal electron acceptor which under aerobic conditions is oxygen. The use of oxygen as a terminal electron acceptor is favoured by *E. coli* due to a high energy yield. When oxygen is absent *E. coli* expresses reductase enzymes which allow it to use alternative terminal electron such as nitrate, tetrathionate or propanediol. Anaerobic respiration in *E. coli* liberates less energy as the bacteria is unable to fully oxidize substrates, consequently in the human gut where oxygen is limited anaerobic bacterial species of the Bacteroidetes and Firmicutes phyla dominate whilst *E. coli* is present in low abundance [43]. However during gastroenteritis this composition changes with *E. coli* and other Enterobacteriaceae expanding significantly [88]. An increased abundance of *E. coli* has been observed

in numerous chronic and acute inflammatory conditions [35, 141, 160, 154]. Anaerobic metabolic pathways play an important role in the expansion of *E. coli* during intestinal inflammation.

Enzymes involved in anaerobic metabolism in *E. coli*, such as reductases, rely on a molybdenum co-factor synthesised by the enzyme MoaA, deletion of this gene prevents *E. coli* from performing anaerobic respiration [162]. Deletion of *moaA* significantly reduces the growth of *E. coli* in an inflamed mouse intestine [162]. Additionally deletion of nitrate reductase genes impairs the growth of *E. coli* in the inflamed gut as it is no longer able to utilise nitrate as a terminal electron acceptor [162, 141]. The nitrate reductase genes present in *E. coli* are homologous to those in *Salmonella enterica* serovar Typhimurium which is a closely related enteric pathogen that causes gastroenteritis in order to proliferate. *S.* Typhimurium is known to be an effective coloniser of a broad range of hosts, producing a characteristic growth bloom in the gut of a newly colonised individual alongside gastroenteristis. How *S.* Typhimurium achieves such efficient growth in the presence of a native microbiota has been the subject of numerous studies which have concluded that it exploits the host inflammatory response.

1.9.1 Colonisation by *S.* Typhimurium

Upon entry into the host *S.* Typhimurium is faced with the challenge of surviving and proliferating in an environment which is already occupied by a multitude of commensal microbes. To establish itself within the new host *S.* Typhimurium creates a new environment which it is genetically equipped to exploit. First it deploys its T3SS to invade the host intestinal epithelium, its presence is detected by Pattern Recognition Receptors (PRRs) that are present on multiple host cell types. These receptors initiate a signalling cascade leading to the activation of immune cells. Release of cytokines (key immune signalling molecules) promote the recruitment of immune effector cells to the intestine. One of the first cell types recruited are

neutrophils which migrate into the intestinal mucosa and even into the intestinal lumen and can be detected in the *S.* Typhimurium induced diarrhoea. Neutrophils, and other immune cells, attempt to clear the infection by producing both Reactive Oxygen Speciess (ROSs) and Reactive Nitrogen Speciess (RNSs). While ROS and RNS are effective at killing bacteria they also react with compounds in the gut lumen producing a new variety of metabolites not normally present [125, 132]. One such compound is tetrathionate, produced when ROS reacts with thiosulfate present in the gut lumen. *S.* Typhimurium possess the *ttr* operon encoding a set of genes that allow the bacteria to utilise tetrathionate as a terminal electron acceptor for anaerobic respiration. Native commensals are incapable of utilising tetrathionate and so *S.* Typhimurium faces no competition for this metabolite and can therefore achieve a growth advantage through tetrathionate utilisation. *S.* Typhimurium lacking the *ttrA* gene are unable to metabolise tetrathionate and have reduced fitness in a mouse model of infection [161], and are unable to outcompete commensals.

Tetrathionate is not the only metabolite utilised by *S.* Typhimurium to outcompete commensals, it is also capable of respiring propanediol. Deletion of the *pdu* operon responsible for propanediol utilisation, decreases the competitive fitness of *S.* Typhimurium compared to wild type [48]. Additionally *S.* Typhimurium is capable of utilising nitrate derived from RNS as deletion of the nitrate reductase genes *napABC*, or their regulator *narP*, imposes a fitness cost in mouse models [85]. Therefore *S.* Typhimurium is genetically equipped to use a variety of inflammatory by-products, such as tetrathionate, propanediol and nitrate, allowing it to out-compete commensals and proliferate in the gut lumen. Increased proliferation in the gut combined with gastroenteritis is proposed to promote onward transmission to new hosts.

1.9.2 Importance of Inflammation for Growth of Both *S.* Typhimurium and *E. coli*

An inflammatory response from the host is essential for the growth bloom of *S.* Typhimurium. Deletion of T3SS genes in *S.* Typhimurium prevent it from stimulating an inflammatory response [161, 48]. *TtrA* and *pdu* deletion mutants exhibit reduced growth compared to wild-type, however when these mutants are combined with a T3SS deletion there is no apparent difference. Specifically a *ttrA* T3SS double mutant grows to equivalent levels of a T3SS single mutant [161, 48]. Alternatively, if the host immune response is defective both *ttrA* mutant and wild-type *S.* Typhimurium grow to equivalent levels, although growth is reduced compared to mice with a functional immune system. Deletion of the host *cybb* gene prevents the immune system from generating ROS, consequently no tetrathionate is generated and wild type *S.* Typhimurium are unable to achieve an advantage over the mutant strain [161]. The host immune response also provides a source of nitrate, specifically the host *Nos2* gene encodes Inducible Nitric Oxide Synthetase (iNOS) which generates RNS. *S.* Typhimurium infected mice display elevated levels of nitrate in their colon compared to mock treated mice suggesting that the bacteria is promoting nitrate production [85]. *S.* Typhimurium can then use the host derived nitrate for anaerobic respiration. The nitrate reductase enzymes in *S.* Typhimurium responsible for this are shared with *E. coli*. Use of molecular inhibitors which prevent the formation of RNS by host enzymes impairs the growth of *E. coli* in a mouse model [162]. Additionally, deletion of the host genes responsible for RNS generation reproduces the same impairment of *E. coli* growth [141]. Therefore the growth of both *S.* Typhimurium and *E. coli* is heavily influenced by the host inflammatory response.

The host immune system uses small proteins called cytokines to transmit signals between cells, virulent *E. coli* is capable of detecting these peptides and altering gene expression in response. Specifically the cytokines TNFα, IL1-β, IL-6, IL8 or IFNγ all resulted in increased expression of metabolic and virulence genes in the ExPEC reference strain CFT073 [44,

122]. The presence of these cytokines stimulated an increased rate of growth of CFT073. It is unlikely that the *E. coli* is directly metabolising these cytokines as the concentrations used are in the nanogram range and the total growth achieved by the strain is not increased, rather the rate at which stationary phase was reached was increased [44, 122]. Moreover bacterial binding to the cytokines could be saturated, inconsistent with continual metabolic processing. The host inflammatory response can be detected by virulent *E. coli* and provide them with a source of metabolites from which to derive energy.

1.9.3 Colonisation by MDR ExPEC

International travel is contributing to the global spread of antibiotic resistance with evidence indicating that ESBL Enterobacteriaceae, including *E. coli*, are capable of colonising healthy individuals in the absence of antibiotic selection. Numerous studies have demonstrated that international travellers acquire ESBL producing Enterobacteriaceae and are still colonised upon their return. The imported bacteria can then spread further within the community with household transmission of imported strains being observed [7]. Rates of colonisation vary but are typically around 30% [7, 113, 109] although some have reported rates as high as 70% [120]. Of the colonising ESBL bacteria *E. coli* were observed to be the most frequent [144], particularly those carrying CTX-M [144, 7, 75]. Whilst exposure to antibiotics was identified as a risk factor for colonisation, region of travel also posed a high risk, particularly Asia or South East Asia [7, 144, 113, 109, 12, 164, 128]. In addition travel to Asia was also associated with a longer duration of colonisation [128]. Geographic region and antibiotic exposure were not the only risk factors identified, gastroenteritis and chronic inflammatory bowel disease were also identified as a risk factors for colonisation [164, 109, 144, 7]. This suggests that individuals experiencing gastointestinal inflammation were pre-disposed to colonisation by MDR bacteria. Furthermore individuals colonised by an ESBL *E. coli* frequently reported experiencing traveller's diarrhoea [76], despite the invading *E. coli* being predominantly Ex-

PEC rather than InPEC. Colonisation by an ESBL bacteria was also associated with an expansion of the Proteobacteria phyla (which includes Enterobacteriaceae), that are known to proliferate under inflammatory conditions [12, 82]. Data from international travllers indicates that healthy inviduals are readily colonised by ESBL producing bacteria, with CTX-M producing *E. coli* being the most common. Gastroenteritis is associated with colonisation; individuals with chronic bowel disease are at a higher risk for colonisation and those who are colonised frequently report experiencing traveller's diarrhoea. Surprisingly colonisation by an ESBL bacteria results in non-significant changes to the microbiome composition[12]. This suggests that commensal strains are displaced by their MDR counterparts. ESBL producing *E. coli* are therefore effective colonisers of the human gut being able to colonise healthy travellers as well as transmit within households [7, 98]. ESBL *E. coli* are able to compete with native microbes in the absence of antibiotics and likely displace resident commensal *E. coli*. Colonisation by ESBL producing bacteria is associated with gastroenteritis suggesting that ESBL *E. coli* have evolved to exploit the host inflammatory response.

1.10 Colonisation and Evolution of Resistance

Exploitation of the host inflammatory response can facilitate host colonisation by MDR ExPEC clones. Increased colonisation efficiency also promotes transmission between hosts. The ability for bacteria to colonise and transmit between individuals has important implications for the evolution of antibiotic resistance. Prolonged colonisation increases the likelihood of exposure to antibiotics, thereby promoting aquisition of resistance, while increased transmission promotes the spread of resistance genes. Mathematical modelling has demonstrated a link between duration of colonisation and AMR with antibiotic resistance genes predicted to provide a fitness advantage to bacteria which persist in hosts whilst incurring a fitness cost in those which are only transient colonisers [81]. Importantly the predictions from this model have been verified using *S. peumonieae*, the prevalence of resistance in

specific serotypes was associated with duration of carriage [81]. *E. coli* ST131 has been observed to display prolonged colonisation in individuals compared to other ExPEC lineages and when compared to other ESBL producing *E. coli* [111], with a calculated half-life of 13 months. Whilst this observation was made in a long term care facility, extended colonisation can be observed in healthy individuals returning from a period of international travel. For example Tangen et al. observed that 5 out of 21 (23.8%) individuals were still colonised by ESBL producing bacteria 6 months after their return from abroad [144]. Paltangsing et al. observed a similar duration of carriage with 26 out of 113 (23%) individuals still colonised at 6 months whilst Arcilla et al. observed that 65 out of 577 (11.3%) were still colonised after one year [113, 7]. Travel to Asia was associated with an increased risk of ESBL colonisation but also a longer duration of colonisation [128]. This suggests that ST131, and ESBL producing Enterobacteriaceae from Asia, colonise individuals for longer than other ExPEC and are therefore predicted to benefit from AMR. Understanding how ST131 and other MDR ExPEC colonise individuals is therefore of great interest.

1.11 Aims and Experimental Approach

E. coli ST131 is a highly successful MDR clone that has spread globally and now contributes significantly to the burden of MDR infections. There is evidence demonstrating that ST131 is an efficient coloniser of new hosts being imported by international travellers and spreading within households. AMR surveillance data has identified a number of other emerging MDR ExPEC clones. Determining the genetic factors permitting the success of ST131 and other clones has identified that genes involved in anaerobic metabolism are undergoing selective pressure in the ST131 lineage. Anaerobic metabolism is important for the growth of *E. coli* during gastroenteritis, which is a frequently reported symptom by travellers who are colonised by ESBL producing bacteria. Investigations into *Salmonella* colonisation mechanisms has identified that it has evolved to exploit the host inflammatory response as a source

of metabolites for anaerobic respiration. **This leads us to hypothesise that: *E. coli* ST131 is exploiting the host inflammatory response to transmit between hosts, in a manner similar to that observed in *Salmonella*. We further hypothesise that this phenotype may be shared with other MDR ExPEC lineages which are emerging as pandemic clones.**

To test these hypotheses we will:

1. Use the previously published pangenome methodology to determine biological processes experiencing selective pressures in a selection of *E. coli* lineages. There will be a focus primarily on lineages which are dominant ExPEC but will also include other pathotypes such as EHEC and EPEC.

2. To test whether MDR *E. coli* strains are exploiting the host inflammatory response to colonise hosts we will perform *in vitro* competition assays. These assays will utilise cell culture models to mimic the intestinal environment and assess whether an MDR ST131 strain is able to outgrow a commensal *E. coli*. The inflammatory response that these two strain elicit from the human cells will also be measured.

3. We will use an *in vivo* mouse model to monitor the ability of MDR and non-MDR *E. coli* to colonise hosts in the absence of antibiotic selection.

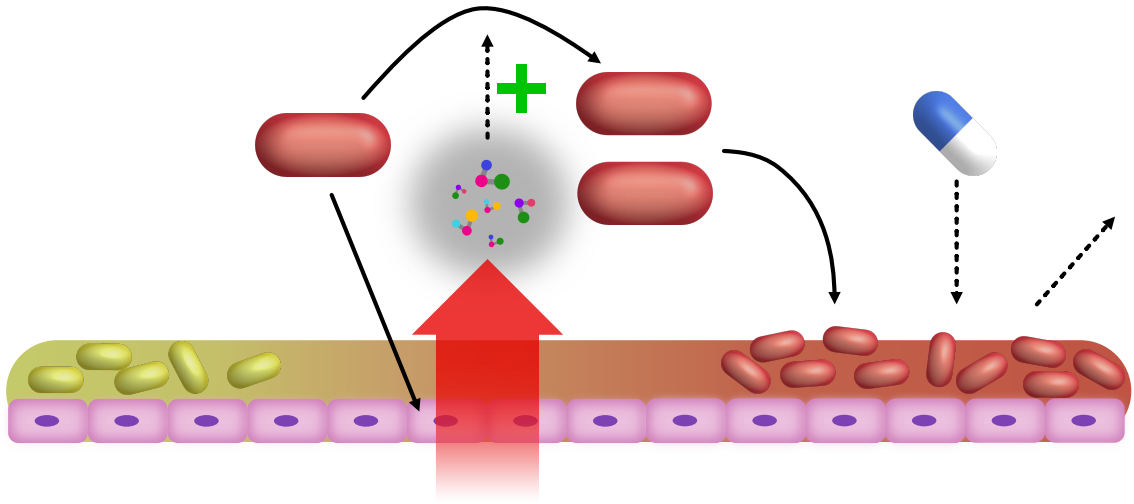

Fig. 1.4 Schematic of thesis hypothesis. Invading MDR *E. coli* (red) stimulate an inflammatory response from the host (red arrow). The host response generates a variety of new metabolites (grey cloud) that allow the invading pathogen to outcompete resident commensals (green). Successful host colonisation can faciliate exposure to antibiotics (white and blue pill) as well as onward transmission to new hosts.

1.11.1 Wider Importance

Understanding the transmission of ST131 will allow us to develop new strategies to combat the spread of AMR. Specifically, developing mechanisms to prevent colonisation by MDR *E. coli* thereby preventing the transit of AMR genes around the globe. For example it has already been demonstrated that alterations in diet can dramatically affect the growth of *Salmonella* and *E. coli* in humans [165], specifically increased consumption of fat with decreased fibre increases the growth of both *E. coli* and *Salmonella* 10-fold. In addition certain strains of *E. coli* can prevent the growth of *Salmonella* in the human gut [165]. A greater understanding of how MDR bacteria colonise hosts can lead to the development of targeted approaches to de-colonise or prevent MDR colonisation.

Chapter 2

Materials & Methods

2.1 *E. coli* Dataset Assembly

Pertinent to Chapter 3.

2.1.1 Selection of relevant *E. coli* lineages

In order for our analysis to be robust it is essential to select the correct lineages for analysis. Specifically multiple highly resistant lineages are required so that any observations are not unique to specific lineages but rather shared amongst all resistant populations. Moreover it is important to have multiple comparator lineages that share pathogenic characteristics with MDR lineages but lack antibiotic resistance genes. Multiple STs which are frequently identified as MDR were chosen, such as ST131, ST69 and ST167 [32, 116, 20, 169]. ExPEC lineages which are not typically MDR uch as ST73 and ST95 were selected as relevant comparators [32]. Additional pathogenic phenotypes were included, specifically the enterohaemorrhagic *E. coli* (EHEC, ST17 and ST21) including the O157 serotype (ST11), as well as Enteropathogenic *E. coli* (EPEC, ST3 and ST28). The generalist and commensal lineage

Table 2.1 Enterobase Assembly Criteria for *E. coli*

Number of bases	3.7 Mbp to 6.4 Mbp
N50 Value	>20 kb
Number of contigs	$\leq$ 800
Proportion of scaffolding placeholders (N's)	<3%
Species assignment using Kraken	>70% contigs are assigned

ST10 was also included. In total 20 lineages were chosen: 3, 10, 11, 12, 14, 17, 21, 28, 38, 69, 73, 95, 117, 127, 131, 141, 144, 167, 372 and 648.

2.1.2 23,567 genome assemblies downloaded from Enterobase

Genomes from the selected lineages were downloaded from Enterobase which is a free publicly available database of bacterial genomes hosted at Warwick [171]. Their database is constructed by pulling reads from publicly available sequencing repositories such as NCBI, SRA and ENA. The reads are then put through an assembly pipeline to generate genome assemblies. Briefly, the reads are trimmed with Sickle before assembly using the genome assembler SPAdes. The raw reads are then mapped onto the resulting assembly and misassembly errors are corrected. The assembly is then put through their quality assessment pipeline which attempts to removes assemblies which have significant contamination, as detected by Kraken. The final assembly must pass several quality criteria before it is released into the MLST database (Table 2.1). The Enterobase database was queried using their online platform to identify genome assemblies assigned to the Warwick ST groups chosen (Section 2.1.1). The genome assemblies along with associated metadata were downloaded using a python script [https://github.com/C-Connor/EnterobaseGenomeAssemblyDownload]. For the 20 selected lineages a total of 23,567 genomes were downloaded (Figure 2.1).

2.1.3 Removal of Duplicate Genome Assemblies

Due to the wide and varied sources of sequence data that have contributed to Enterobase it is entirely possible that identical isolates have been sequenced multiple times. To avoid influencing the results the dataset was first de-replicated using Mash 1.1.1 and a custom R script [108] [https://github.com/C-Connor/MashDistDeReplication]. The Mash programme greatly reduces the complexity of genomes by condensing them into short k-mer 'sketches' that are passed through a hash function. It is then feasible to compute pair-wise comparisons of the hashes for all the genomes, producing a distance matrix for all the isolates. This distance matrix was then analysed using the custom R script which clusters genomes with 0 mash distance and selects a single representative genome from each cluster to keep. While the Mash sketching process is more feasible than whole genome comparisons it was still too time consuming to run on the total dataset, instead the de-replication pipeline was run on individual lineages. The ST10 population had the largest number of genomes removed with just under half of the samples removed (1720 genomes removed, 2370 remaining, Figure 2.1). Further examining the metadata for this population it became apparent that there were several time course evolution experiments that contributed to the number of replicated assemblies. These experiments are apparent on the phylogenetic tree (Figure 2.2) as a cluster of 0 length branches. Examining the reference for these samples showed that these were experiments where a K12 lab strain was exposed to various stimuli and repeatedly sequenced to look for SNP changes [102, 83]. Whilst only two of these types of experiments have been highlighted in the tree there are likely more. Following the de-replication pipeline these clusters of 0 length branches are absent from the dataset (Figure 2.3).

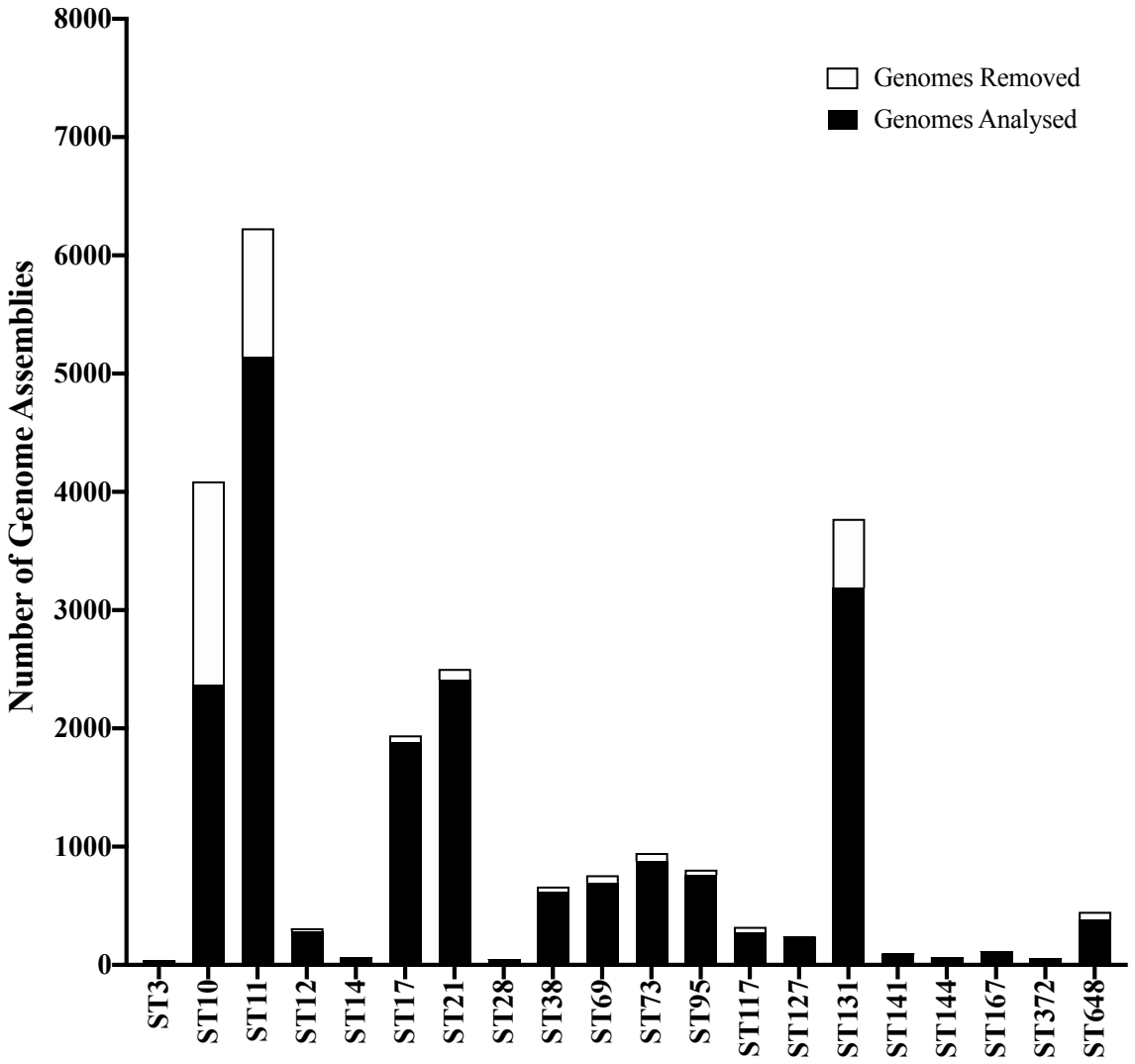

Fig. 2.1 Number of genome assemblies downloaded and those remaining post de-replication for each sequence type. Total height of bars indicates the total number of genome assemblies downloaded. Black shaded bars indicate the number of assemlies remaining after filtering. White bars indicate number of genomes removed.

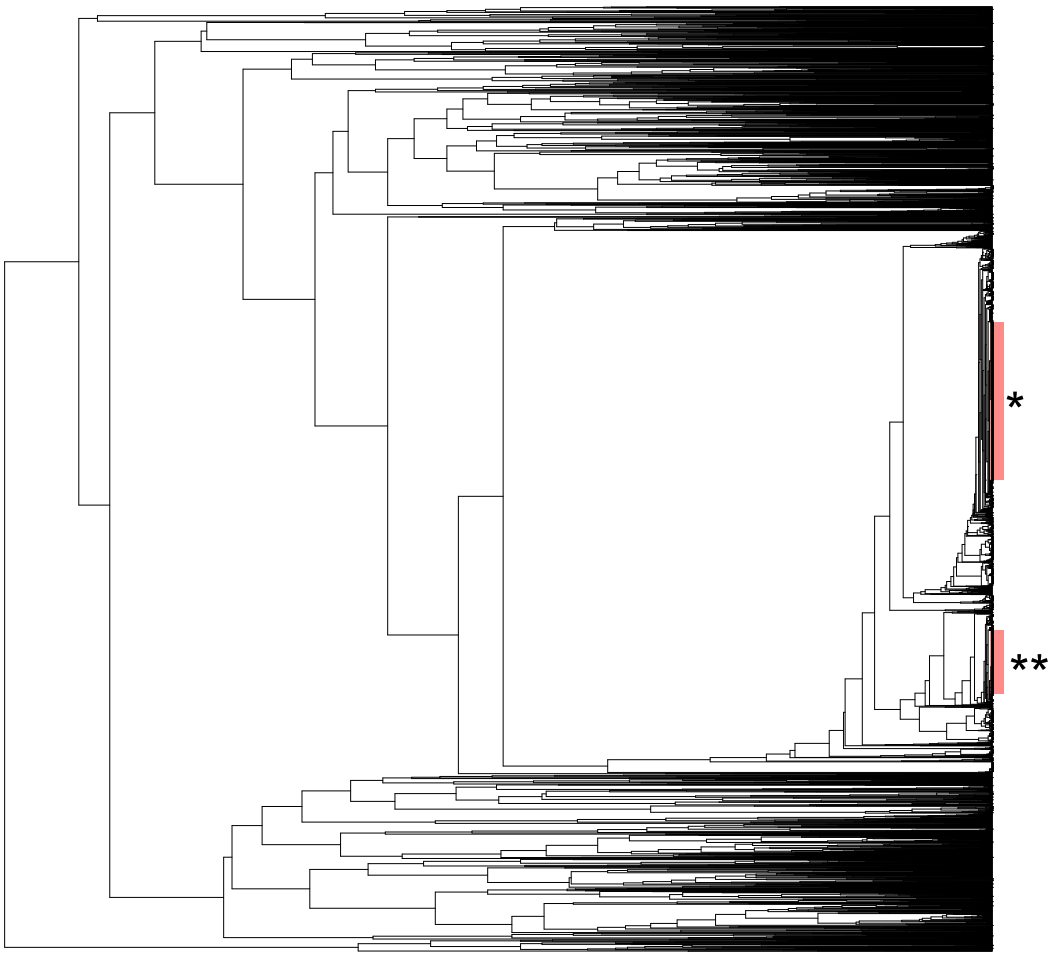

Fig. 2.2 Phylogenetic tree of the ST10 lineage before removing duplicate genomes. Identical genomes appear as branches with 0 length adjacent to each other in the tree. Two selected clusters of identical genomes are highlighted in red. Regions marked with [*] were from [102] whilst those marked with [**] are from [83]; both of these are experimental evolution studies that employ repeated sequencing of the same bacterium experiencing a selective pressure.

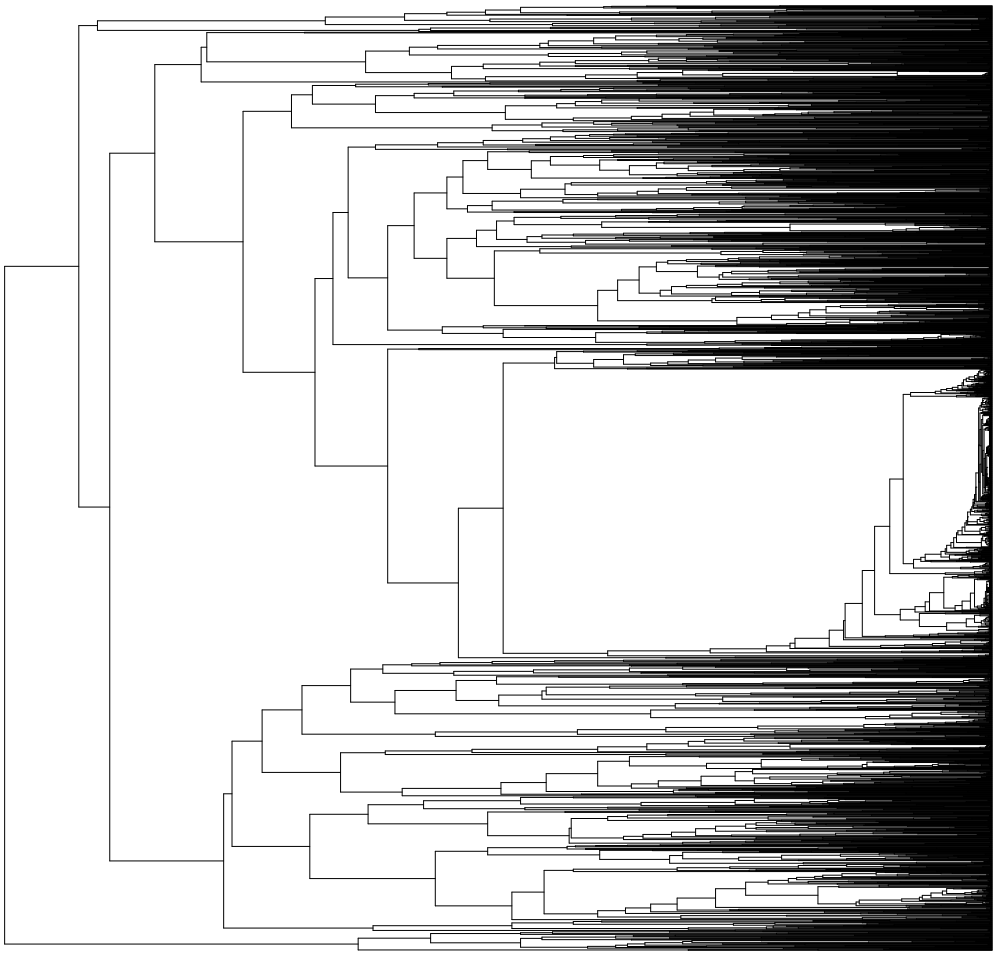

Fig. 2.3 Phylogenetic tree of the ST10 lineage post filtering. Large clusters of 0 length branches are absent.

2.1.4 Removal of Genome Outliers

The de-replication pipeline will identify closely related genomes but is unable to account for highly divergent assemblies, such as from an erroneous sequence type assignment. Whilst Enterobase employs stringent quality filters when adding new genomes to the database there is no manual verification. To account for this the phylogenetic trees for each ST population were visually assessed for outlying assemblies. The phylogenetic trees were produced using MashTree 0.36.2 which also utilises the Mash program to generate trees for large datasets [78]. Outlying samples were identified by being placed outside the main population cluster on a single branch. The phylogenetic trees were manually pruned until the outermost assembly was part of a larger cluster (Figure 2.4). Using this pruning approach a total of 4 genome assemblies were removed from the dataset, one from ST11, one from ST21 and two from ST127.

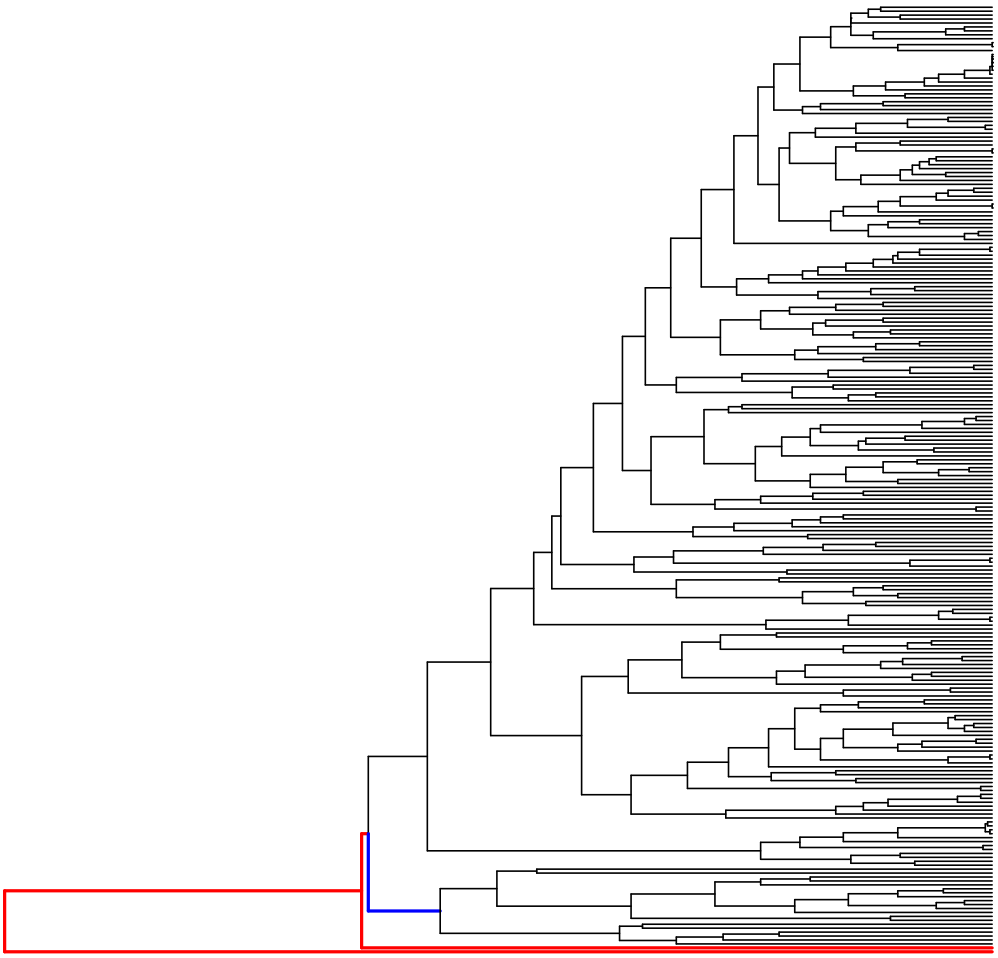

Fig. 2.4 Phylogenetic tree of the ST127 lineage after duplicate genomes have been removed. Outlying genomes, highlighted in red, will be removed until the next outer branch consists of multiple genomes, highlighted in blue. A total of 4 genomes were removed from the dataset using this method.

2.1.5 Genome Annotation, AMR Identification, and Genome Metrics

Genomes which passed through the de-replication and phylogenetic pruning were annotated using Prokka 1.12 [137]. AMR gene content in individual genome assemblies was assessed using Abricate 0.8 [https://github.com/tseemann/abricate] and the Resfinder database (March 2018) [17]. The number of CDS annotated was calculated by counting the number of features annotated with 'CDS' in the GFF annotation file produced by Prokka. Genome length was determined from the metadata downloaded from Enterobase.

2.1.6 Pangenome Assembly

There are multiple variables that can be controlled when constructing a pangenome, one of which is the gene identity threshold. The pangenomes presented here were generated with Roary 3.10.2 [112] using the GFF annotation files produced by Prokka , pangenomes were generated with and without the paralog splitting function. An identity threshold of 95% was used meaning that if the protein sequence of two genes differs by more than 5% then they will be called as separate genes and given separate entries in the pangenome output. This is a highly conservative identity cut-off and will therefore separate alleles of the same gene causing that gene to be over-represented in the pangenome. This has the additional effect of reducing the measured frequency that the gene occurs, moving the gene into the accessory compartment. It is then possible to query the accessory pangenome for over-represented genes and thereby infer which genes are exhibiting variation across the population. An increase in variation for a particular biological function in a bacterial population is a strong indicator of a selective process acting on that function. By assigning genes a functional role we can determine which biological processes are exhibiting increased variation, and inferred selection.

2.1.7 Functional Pangenome Annotation and GO Enrichment Analysis

The nucleotide sequences in the pangenome fasta file from Roary 3.10.2 were translated into peptide sequences using a custom python script. Functional annotation of the resulting peptide pangenome sequences was performed with emapper-1.0.3-3-g3e22728 [65] based on eggNOG orthology data [66]. Sequence searches were performed using DIAMOND [21]. The COG categories assigned from the functional annotation were summarised in R and plotted against the proportion of AMR genes present in lineages. Significance of correlations was determined using a linear regression model. The functional annotation also assigned Gene Ontolotgy (GO) terms to the pangenome genes. GO term enrichment analysis was performed using the topGO package [1] in R. GO terms which were enriched in the combined accessory pangenome of the 5 lineages with the highest AMR burden (ST38, ST69, ST131, ST167 and ST648) compared to the other *E. coli* linages. The same approach was used to test for enrichment of GO terms in the core genome. Statistical tests in topGO were performed using the 'weight01' algorithm with Fisher's exact test.

2.1.8 Phage Burden, IGR Variation and Recombination

Phage burden was estimated by comparing pangenome sequences to the prophage and virus database from PHAST [170] using BlastX [22]. The resulting hits were filtered to remove matches with less than 80% sequence identity or less than 20 residues in length. To test the contribution that phage elements made to the observed associations between the pangenomes and AMR carriage, phage elements were removed and the association analysis was repeated. Inter-genic region variation was assessed using Piggy 1.2 with default parameters [148]. The resulting output was processed and analysed in R. Recombination was predicted using Gubbins 2.4.1 [31]. Whole genome alignments for Gubbins were produced using Ska 1.0 [60] which converts genome assemblies into k-mers with the 'ska fasta' command, these k-mers were then mapped to a randomly selected reference genome using 'ska map'. For

ST131 a genome was randomly selected from clade C, for all other lineages the genome was randomly selected from the entire lineage. The resulting alignment was then used as input for Gubbins with the hybrid tree flag, all other options were default. Results were visualised in Phandango [57].

2.2 Eukaryotic and Bacterial Cell Culture

Pertinent to Chapter 4.

2.2.1 Bacterial Strains and Antibiotics

For details of bacterial strains used please see Table 2.2. For details of antibiotic compounds used please see Table 2.3.

2.2.2 Bacterial Culture

Bacterial growth media was prepared from powder stocks using distilled water following the manufacturer's recommendations. Media was sterilised by autoclaving (121°C for 15 mins) before use, where necessary antibiotics were added to cooled media. Bacteria were grown on Luria Bertani (LB) agar (Miller, EO Labs KM0114) at 37°C or in liquid LB broth (Miller, VWR Chemicals 84649.0500) at 37°C with agitation at 200rpm. Stocks of bacteria were stored in -80°C in LB broth +10% glycerol (Sigma G629-500ML).

Table 2.2 Bacterial Strains

Strain	Phylogroup	MLST	Origin	Source
K12 MG1655	A	ST10		
K12 MG1655 mCherry	A	ST10		[18]
811E7	A	ST10	Healthy human, gut	P. Hawkey
F022	B2	ST131 A	Bacteraemia	[42]
F037	B2	ST131 A	Bacteraemia	[42]
F048	B2	ST131 B	Bacteraemia	[42]
F054	B2	ST131 B	Bacteraemia	[42]
F047	B2	ST131 C	Bacteraemia	[42]
F104	B2	ST131 C	Bacteraemia	[42]
F016	B2	ST131 C	Bacteraemia	[96]
F084	B2	ST73	Bacteraemia	[2]

Table 2.3 Antibiotic Compounds

Antibiotic	CAS	Form	Concentration	Solvent
Ciprofloxacin (Alfa Aesar J61317)	85721-33-1	Powder	1μg/ml	Water
Sulfamethoxazole (Sigma S7507-100G)	723-46-6	Powder	512μg/ml	Acetone
Ampicillin (Roche 10835242001)	69-53-4	Powder	100μg/ml	Water
Hygromycin B (Invitrogen 10687010)	31282-04-9	Liquid (PBS, 50μg/ml)	300μg/ml	PBS

Table 2.4 Plasmids

Plasmid	Tag	Resistance	Source
pMN402	GFP	Hygromycin B	[6]
pFPV25.1	GFP	Ampicillin	Addgene
pCherry8	mCherry	Hygromycin B	Addgene

2.2.3 Plasmids and Transformation

Plasmids were transformed into strains using electroporation. Cells were made competent by growing culture to mid-log phase before washing in ice cold 10% glycerol three times. Plasmid DNA was added to competent cells before electroporating at 1.8kV using an Eppendorf ePorator. Immediately after electroporation pre-warmed (37°C) SOC media (Sigma S1797-10X5ML) was added to cells and bacteria were allowed to recover for 1 hour at 37°C with agitation at 200rpm, before plating onto selective agar.

2.2.4 Eukaryotic Cell Culture

Caco-2 (ATCC HTB-37) and INT407 (ATCC CCL6) were obtained from Xuan Wan Kang (Piddock group). Cells were grown in Dulbecco's Modified Eagle's Medium – high glucose (DMEM) (Sigma D596-500ML) supplemented with 10% Foetal Bovine Serum (Gibco 10500-064) and 25mM HEPES (Gibco 15630-056) at 37°C 5% CO_2. Cells were sub-cultured by washing with sterile Dulbecco's Phosphate Buffered Saline (PBS, Sigma D8537-500ML) before treating cells with Trypsin - EDTA (Sigma T3924-100ML) for 5 minutes, cells could then be detached by vigorously resuspending with fresh supplemented DMEM. Cells number was calculated using a haemocytometer (Hirschman 8100103), cells were transferred to new flasks at a target density of 1×10^4 cells per cm^2. Media was renewed every 3 to 4 days. Cell stocks were maintained by transferring to new flasks when cells were at 80-90% confluency. Cell stocks were maintained in T75 flasks (Corning 430641U) whilst experiments were conducted in 12 well (Corning 3513) or 24 well plates (Corning 3526). Cells were

not used past passage 15. Caco-2 cells were allowed to spontaneously differentiate over a 2 week period after reaching confluency before use in experiments. INT407 cells were used in experiments once they had reached confluency. Long term cell stocks were stored in liquid nitrogen, cells were preserved in DMEM with 5% DMSO (Sigma D2650-100ML). Cells to be stored were resuspended in preservation media at a concentration of 1×10^6 cells per ml and transferred to cyrotubes (Nalgene 5000-0020), cells were cooled gradually in a Mr Frosty (Nalgene C1562-EA) at -80°C overnight before transfer to liquid nitrogen.

2.2.5 Eukaryotic and Bacterial Cell Co-culture

Infection

Bacteria were grown in LB broth overnight at 37°C with shaking at 200 rpm, where appropriate antibiotics were included to maintain selection for plasmids. The colony forming units (CFU) of the overnight culture was calculated by plating serial dilutions onto LB agar and enumerating colonies. Eukaryotic cells were cultured in 12 or 24 well plates and grown until they reached confluency. Caco-2 cells were allowed to spontaneously differentiate before experiments. The number of cells present in the monolayers was determined by detaching monolayers with Trypsin / EDTA and counting multiple wells with a haemocytometer. Media on the eukaryotic monolayers were replaced with fresh supplemented DMEM before addition of bacteria. Overnight bacterial cultures were diluted to an appropriate concentration in PBS before adding an appropriate volume to the eukaryotic monolayers. Bacteria were added to the monolayers at a target Multiplicity Of Infection (MOI) of 1. Bacterial and eukaryotic cells were grown together at 37°C 5% CO_2 statically. Co-cultures were used for a variety of experiments detailed below.

Fluorescent Growth Curves

Bacterial and eukaryotic cells were cultured as previously described. Fluorescence and optical density measurements were recorded using a Tecan Spark plate reader (Tecan) which maintained a constant temperature of 37°C with 5% CO_2. To avoid disruption of the eukaryotic monolayer cells were cultured statically. Optical readings were taken every 20 minutes, GFP was measured with excitation 485/20; emission 535/25; gain 100, mCherry was measured with excitation 580/20; emission 635/35; gain 125. Data was normalised to the reading taken at 20 minutes to compensate for media equilibration.

Microscopy

Bacterial and eukaryotic cells were cultured as previously described. Cells were infected in 12 well plates. Cells were imaged on a Nikon Ti Inverted widefield fluorescence microscope with a 20x phase contrast objective lens. Cells were imaged in a chamber heated to 37°C with 5% CO_2 and no agitation. Images were acquired with a brightfield phase contrast channel, GFP and RFP channels. For each culture condition 4 imaging positions were chosen, with 7 z-stacks at 2.5μm spacing. Images were acquired every 10 minutes for 7 hours in the XYZCT order. Acquired images were processed in Fiji [136], to produce a maximum intensity z-projection. Brightness and contrast were adjusted so that fluorescence at the final timepoints was at maximum intensity.

Flow Cytometry

Bacterial and eukaryotic cells were cultured as previously described. After 6 hours of co-culture cells were vigorously resuspended by pipetting to disrupt the eukaryotic monolayer. A 1 in 100 dilution of the culture suspension was measured on an Attune NxT flow cytometer (Invitrogen) until 10,000 bacterial events were collected. Gates were determined through the

use of negative controls. Theshold values for foward scatter were set to 400 and side scatter to 500. Voltage for GFP channel was set to 350 and voltage for RFP set to 350.

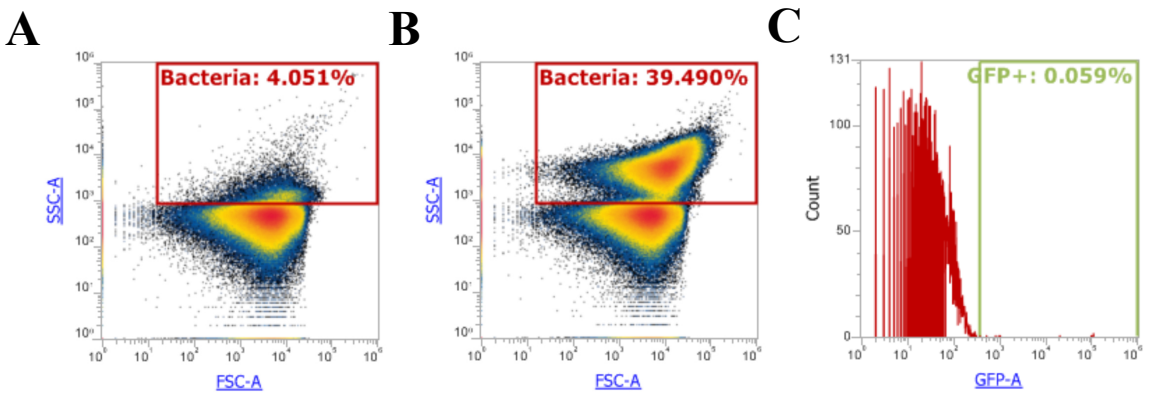

Fig. 2.5 Gating strategy for flow cytometry analysis of bacterial growth. Background events in the absence of bacteria (A) can be distinguished from bacteria (B). Background events do not fluoresce (C).

CFU Determination

Bacterial and eukaryotic cells were cultured as previously described. After 6 hours cell co-cultures were vigorously resuspended by pipetting to disrupt the eukaryotic monolayer. A 20μl aliquot of the suspension was serially 10-fold diluted in sterile PBS. A range of dilutions (1×10^{-3} to 1×10^{-8}) were spot plated (20μl volumes) in triplicate onto LB agar with and without antibiotics. Bacteria were grown on agar overnight at 37°C before enumerating colonies. Dilutions where more than 3 colonies were present and individual colonies could clearly be distinguished were used for counting.

Enzyme Linked Immunosorbent Assay (ELISA)

Co-culture supernatants were collected and centrifuged to remove debris. As a positive control cells were stimulated with 20μg/ml of LPS (*E. coli* O111:B4, Sigma, L3023) for 6 hours before collecting supernatant. Supernatants were stored at -80°C before assaying using the

Multi-analyte ELISArray kit (Qiagen MEH008A, 336161) following manufacturer's instructions. Optical readings were taken using the Tecan Spark plate reader (Tecan) measuring absorbance and 450nm and 570nm. Absorbance readings at 570nm were used to correct for optics by subtracting from the absorbance at 450nm, as recommended by the manufacturer.

2.3 Mouse Models

Pertinent to Chapter 5.

2.3.1 Housing and Colonisation

Germ-free C57BL/6 mice were housed at the International Microbiome Centre at the University of Calgary. Mice were colonised with 1×10^9 CFU of bacteria via oral gavage. Colonisation was monitored by determining the CFU in the faecal pellets. CFU assays were performed using UTI chromogenic agar (Sigma 16636-500G-F) with *E. coli* producing pink colonies. Faecal pellets were homogenised in PBS to a concentration of 100mg/ml. The homogenate was serially 10-fold diluted in PBS before spot plating onto UTI chromogenic agar with and without antibiotics. Plates were incubated at 37°C and bacterial growth was enumerated. Strains were distinguished by their varied resistance profile: ST10 pan-susceptible, ST73 sulfamethoxazole resistant, ST131 ciprofloxacin and sulfamethoxazole resistant.

2.3.2 Cytokine Expression

Sections of mouse gut were preserved in RNAlater Stabilization Solution (Thermo Fisher Scientific AM7020) at 4°C for 16 hours to allow the preservative to penetrate the tissues before transferring to -80°C for long term storage. Tissues were homogenised in 2ml screw top tubes using a 5mm stainless steel bead (Qiagen 69989) with a FastPrep-24 5G bead beater (MP Biomedical). Metal beads were cleaned before use by washing in 70% ethanol, drying

then irradiating with UV light and finally autoclaving. Tissues were pulse homogenised for 1 minute at 6m/sec with 1 minute incubation on ice, this was repeated 3 times. The homogenate was centrifuged to remove debris before the supernatant was processed using the RNeasy mini kit (Qiagen 74104) following the manufacturer's instructions. A DNase treatment step was included as per the manufacturer's protocol using the RNase-Free DNase set (Qiagen 79254). RNA concentration was determined using the Qubit RNA BR Assay kit (ThermoFisher Scientific Q10210) on a Qubit 4 fluorometer (ThermoFisher Scientific).

RNA was converted to total cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems 4368814) with the provided random primers, following the manufacturer's instructions. The resulting cDNA was used for probe based qPCR. Probes were obtained from Integrated DNA Technologies (IDT) using their available pre-designed primers and are listed in Table 2.5. The qPCR reactions were performed using PrimeTime Gene Expression Master Mix (IDT 1055771) to which the provided reference ROX dye was added, as per the manufacturer's instructions. Reactions were performed in a QuantStudio 1 Real-Time PCR System (Applied Biosystems) using a 96-well 0.2ml block. Cycle parameters were: 3 minutes 95°C, 45 cycles of 15 seconds at 95°C, 1 minute at 60°C. Fluoresence was measured at each cycle during the 60°C incubation. Reactions were conducted in 20μl volumes in duplicate with cycle threhsold (Ct) values averaged before normalising to *Pol2ra* and presented as ΔCt values. Negative controls were included: non-reverse transcribed (NRT) in which the reverse transcriptase enzyme was omitted from the reverese transcription reactions, and no-template control (NTC) in which no cDNA was added to the qPCR reaction. Significance was determined using 2-way ANOVA with Tukey's multiple correction.

Table 2.5 qPCR Probes

Assay ID	Gene Query	Species	Ref Seq #	Transcripts Hit	Exon Location
Mm.PT.58.9981538	Il8	Mus_musculus	NM_011339	NM_011339	2 - 4
Mm.PT.58.10005566	Il6	Mus_musculus	NM_031168	NM_031168	4 - 5
Mm.PT.58.41616450	Il1b	Mus_musculus	NM_008361	NM_008361	3 - 4
Mm.PT.58.41769240	Ifng	Mus_musculus	NM_008337	NM_008337	1 - 2
Mm.PT.58.11254750	Tgfb1	Mus_musculus	NM_011577	NM_011577	1 - 2
Mm.PT.58.32778767	Il1a	Mus_musculus	NM_010554	NM_010554	4 - 5
Mm.PT.58.12022572	Il33	Mus_musculus	NM_001164724	NM_133775,NM_001164724	7 - 8
Mm.PT.58.6531092	Il17a	Mus_musculus	NM_010552	NM_010552	2 - 3
Mm.PT.58.13531087	Il10	Mus_musculus	NM_010548	NM_010548	3 - 5
Mm.PT.58.10594618.g	Il23a	Mus_musculus	NM_031252	NM_031252	2 - 3
Mm.PT.58.44003402.gs	S100a8	Mus_musculus	NM_013650	NM_013650	2 - 3
Mm.PT.58.42076891	Cxcl1	Mus_musculus	NM_008176	NM_008176	2 - 4
Mm.PT.39a.22214849	Polr2a	Mus_musculus	NM_009089	NM_009089	22 - 23
Mm.PT.58.12575861	Tnf	Mus_musculus	NM_013693	NM_013693	2 - 4

2.3.3 Histology

Sections of mouse gut were collected and 'swiss-rolled' into histology cassettes. Briefly, gut sections were cleaned by flushing through with PBS, sections were then cut longitudinally before rolling from the proximal end to the distal. Tissues were fixed in 10% formalin for 24 hours before dehydrating. Tissues were dehydrated by incubating in sequentially increasing concentrations of ethanol (70%, 80%, 90%, 95%, 100%) before incubation in Neo-Clear (Merck 109843). After dehydration tissues were incubated in molten paraffin to allow it to permeate the tissue sections. Tissue rolls were then embedded in paraffin blocks such that sectioning would provide a cross section across the length of the gut. Tissue blocks were sectioned into 5μm slices and mounted onto glass slides. Mounted sections were stained with haematoxylin and eosin (H&E) using a Leica ST5020 Multistainer (Leica). Stained sections were imaged using the Axioscan 7 Slide Scanner (Zeiss). Tissue sections were scored using the blind scoring plugin for Fiji [https://imagej.net/plugins/blind-analysis-tools] according to guidelines in [45].

2.3.4 Bacterial Dissemination

At the end of each experiment mouse extra-intestinal tissues (mesenteric lymph nodes, spleen, liver, brain, lung) were collected into 2ml tubes containing 1ml of sterile PBS with a 5mm stainless steel bead. Tissues were homogenised in a TissueLyser (Qiagen) at 50Hz for 3 minutes. The resulting homogenate was plated onto UTI chromogenic agar with and without antibiotics. Plates were incubated at 37°C and resulting bacterial growth was enumerated.

Chapter 3

Pangenome Analysis Identifies an Association Between Metabolism and Antibiotic Resistance

3.1 Introduction

The rise of antibiotic resistance in bacteria is of growing concern with numerous bacterial pathogens becoming increasingly unresponsive to antibiotic therapies. Of particular concern are ExPEC which are the most common causative agent of urinary tract infections (UTI) but can also cause severe blood stream infections. It is estimated that MDR *E. coli* account for significant proportion of MDR infections and are an important contributor to mortality [25]. The rise of resistance in ExPEC has been attributed to a specific lineage called ST131 [3, 70]. In 2007 and 2008 two independent research groups both identified that the ST131 lineage accounted for the majority of CTX-M positive isolates within their collection of clinical samples [103, 30], this was the first time that ST131 was identified as a major contributor to AMR in ExPEC. Strikingly the isolates in both reports were from multiple locations around the globe demonstrating that ST131 had spread globally. Since these first

reports ST131 has been identified in numerous studies and is repeatedly named as the major culprit of MDR ExPEC infections [71, 32, 74, 3]. Consequently, ST131 has been the focus of numerous studies attempting to understand how this specific lineage has become a global contributor to MDR infections. Multiple genomic studies have suggested that the acquisition of AMR genes was not the driving force behind the 'success' of this lineage. For example Kallonen et al. observed that both ST73 and ST131 were causing an equivalent number of blood stream infections in the United Kingdom, and while ST131 frequently carried multiple AMR genes the ST73 isolates remained largely susceptible [74]. The authors concluded that MDR does not determine the success of pathogens in this niche as both highly resistant and susceptible lineages are equally successful. This disparity in AMR carriage has been confirmed in other studies which also identify that MDR and susceptible lineages perform equivalently in the same environment [32, 56, 3]. Additional evidence that AMR is not the determinant of ST131's 'success' comes from Bayesian phylogenetic analysis which determined that the ST131 population was expanding before it had acquired AMR genes [11], suggesting that other factors were promoting the growth of this population. The driving force behind ST131's success remains unclear. Pangenomic analysis of the ST131 lineage has suggested that metabolism is an important factor.

Initial analysis produced a pangenome encompassing multiple ExPEC isolates from ST131 and other dominant ExPEC clones. This analysis identified loci that were unique to ST131, the majority of which were hypothetical proteins, other major categories identified were flagellar and metabolic proteins [6]. Subsequent analysis with an increased number of genomes identified a total of 754 loci that were significantly associated with ST131 genomes compared to non-ST131, of these 292 were metabolic loci [96]. More recent pangenome analysis employed a stringent identity threshold of 95% to separate allelic variants in the pangenome. This analysis identified that ST131 possessed a significantly higher number of alleles of anaerobic metabolic genes than ST73 or ST95 [95]. Importantly, the observation that varia-

tion was unique to ST131 and not shared with other major ExPEC clones (ST73 and ST95) rules out the possibility that this is an ExPEC pathogenicity trait [95]. Moreover the identified alleles were restricted to clades B and C of ST131, which is also where carriage of AMR genes is highest. Further implying that the observed variation is associated with AMR gene carriage and not virulence. The authors conclude that increased variation in anaerobic meatbolism implies that there is a selective pressure acting on this biological process [95].

While ST131 is a major contributor to MDR ExPEC infections it is not the only MDR ExPEC lineage. Surveillance studies have identified a handful of other ExPEC clones that are significantly associated with AMR [116]. For example ST167 (alongside ST131) were found to account for the majority of CTX-M-15 positive isolates from food products in Germany [67]. Additionally, ST167 and ST131 were the most frequently identified carbapenem resistant bacteria in hospitals across China [169]. Whilst analysis of carbapenem resistant isolates in Denmark revealed that the majority were ST38, ST69 or ST167 [59]. Genomic analysis of these populations has also identified that anaerobic metabolism genes are exhibiting increased diversity. For example, multiple unique alleles of dehydrogenase enzymes, which are important in anaerobic metabolism, were found in the ST167 lineage when compared to a commensal lineage [172]. Therefore future analysis should not focus solely on a single lineage but encompass multiple major MDR ExPEC lineages to determine whether adaptations in anaerobic metabolism are common to all MDR ExPEC clones. To achieve this I intend to expand the pangenomic analysis used previously to include multiple MDR ExPEC lineages, as well as non-MDR ExPEC, commensals, and other pathogenic *E. coli* such as EPEC and EHEC lineages.

3.2 Results

3.2.1 Dataset Assembly and Properties

Selection of relevant *E. coli* lineages

In order for our analysis to be robust it is essential to select the correct lineages for analysis. Specifically multiple highly resistant lineages are required so that any observations are not unique to a specific lineage but rather shared amongst all resistant populations. Multiple STs which are frequently identified as MDR were chosen, such as ST131, ST167 and ST648 [116, 67]. Moreover it is important to have multiple comparator lineages that share pathogenic characteristics with MDR lineages but lack antibiotic resistance genes. ExPEC lineages which are not associated with multi-drug resistant such as ST73 and ST95 were selected as relevant comparators [32]. Additional pathogenic phenotypes were included, specifically the enterohaemorrhagic *E. coli* (EHEC, ST17 and ST21) including the O157 serotype (ST11), as well as Enteropathogenic *E. coli* (EPEC, ST3 and ST28). The generalist and commensal lineage ST10 was also included, together totalling 20 lineages (Table 3.1).

Genome length and number of annotated CDS are positively correlated

Initial exploratory analysis of the dataset investigated variation in assembly size and coding variation. Assembly length was extracted from the metadata whilst the number of CDS annotations was extracted from genome annotation files produced by Prokka [137]. ST21 had the highest median genome length with 5.56 million base pairs (Mbp) whilst ST10 had the lowest with 4.85 Mbp (Figure 3.1a). The median for the entire dataset is 5.32 Mbp. Examining genome length by phylogroup it appears that lineages from the same phylogroup exhibit broadly similar genome lengths; phylogroup A lineages are present at the lower end of the spectrum whilst phylogroup B1 lineages are present at the upper end of the spectrum. The ST10 lineage exhibits a large variation in genome size compared to all other populations

Table 3.1 Selected populations of *E. coli* for analysis

	Sequence Type	Phylogroup	Genomes Downloaded	Genomes Analysed	Comments
1	ST3	B1	40	40	EPEC
2	ST10	A	4090	2370	Generalist
3	ST11	E	6230	5137	EHEC O157
4	ST12	B2	311	283	ExPEC
5	ST14	B2	63	62	ExPEC
6	ST17	B1	1942	1884	EHEC
7	ST21	B1	2504	2411	EHEC
8	ST28	B2	47	46	EPEC
9	ST38	D	662	617	ExPEC
10	ST69	D	759	696	ExPEC
11	ST73	B2	946	873	ExPEC
12	ST95	B2	805	758	ExPEC
13	ST117	F	322	269	ExPEC
14	ST127	B2	241	232	ExPEC
15	ST131	B2	3772	3186	ExPEC
16	ST141	B2	94	91	ExPEC
17	ST144	B2	65	65	ExPEC
18	ST167	A	117	115	ExPEC
19	ST372	B2	56	54	ExPEC
20	ST648	F	450	382	ExPEC

(inter-quantile range 4.59-5.06 Mbp), the lineage with the next largest variation is ST372 (inter-quantile range 4.79-5.19 Mbp) . The median for all of the populations assayed is notably higher than the length of the reference *E. coli* genome K12 MG1655 (4.64 Mbp). The observations made for the genome length are mirrored by the those observed in the number of Coding Sequences (CDSs) annotated (Figure 3.1b). Again the ST21 lineage has the highest median number of CDS (5372 genes) whilst the lowest is ST10 (4535 genes). The phylogroups continue to group into clusters with phylogroup B1 at the upper end of the spectrum and phylogroup A at the lower end. The majority of the populations assayed are under the median for the entire dataset (5021 genes). The median CDS count for the ST10 lineage now matches that of the reference *E. coli* K12 MG1655 genome (4535 vs. 4566 respectively). ST10 exhibits a broad range of CDS counts however its inter-quantile range is

almost matched by that of ST372 (4268-4767, 4394-4849 respectively). Directly comparing the length of the genome assemblies against the count of annotated CDS confirms the two are positively correlated (Figure 3.1c). It is more evident that phylogroup A genomes cluster at the lower end of the spectrum where the majority have a low genome length and fewer annotated CDS. Phylogroups D and E occupy the mid-region whilst phylogroup B1 have larger genome assemblies and more annotated CDS.

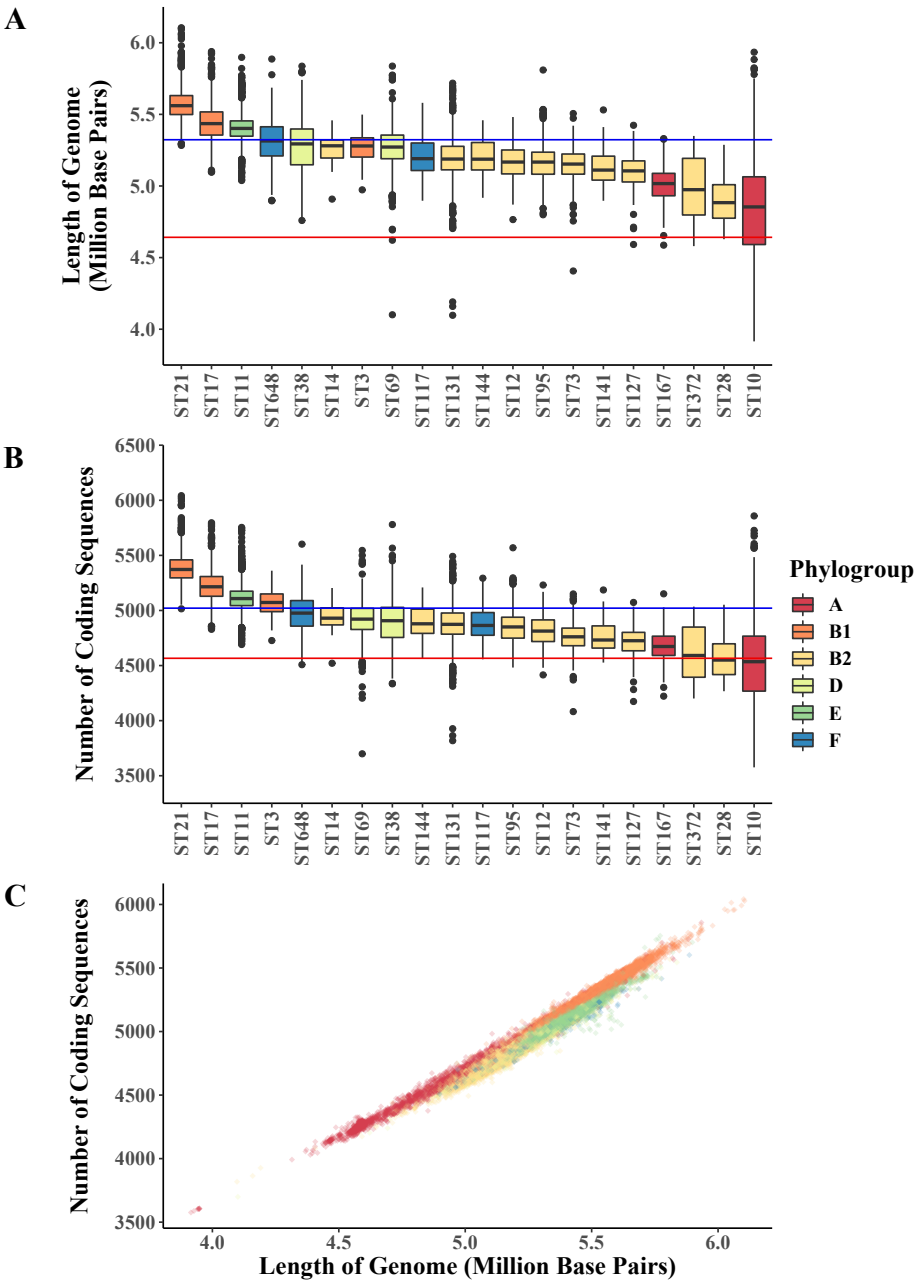

Fig. 3.1 Genome Length and Number of Coding Sequences for each Sequence Type. (A & B) Bold central line within box indicates median for population. The lower and upper bounds for the box indicate the 25th and 75th percentile. Lines indicate value within 1.5x the interquartile range whilst dots indicate values outside that range. Blue horizontal line indicates the average for the entire dataset. Red horizontal line indicates the value of the K12 MG1655 type strain genome. (A, B & C) Boxes and dots are coloured based upon the phyogroup. A) The length of the genome assembly for each ST population. B) The number of annotated coding sequenes for each ST population. C) The number of annotated coding sequences plotted against the length of the genome assembly.

Genome sequences are primarily derived from human sources within the last 20 years

The associated metadata contains information on the source of the sequenced strain. With the exception of ST11 and ST167, the majority of the sequenced strains for the populations had no data available on their source (Table 3.2 & Figure 3.2). While the majority of isolates from ST11 and ST167 came from humans. The second most abundant source for the isolates was human, with the exception of ST117 which had human as the third most abundant source after no data and poultry. Therefore all the populations assayed have a strong bias towards human samples. Information on the location of the sequenced strain was also included in the metadata. Strains in the dataset came from a variety of geographic locations however North American and European strains are over represented in the dataset (Figure 3.3). Again a large proportion of the data does not have location data available (0.357). Lastly, the metadata also contains information on when the sequence data was generated. The dataset also includes a broad temporal range of isolates going back to the 1800's. It became difficult to verify whether these particular strains were genuine. Bacterial genome sequencing grew in practicality and utility around the 1980's, therefore strains with sequencing dates prior to this were treated as missing data. Despite this caveat the majority of strains in the dataset were sequenced in the past 20 years. Overall the dataset captures *E. coli* recently sequenced from across the globe with an emphasis on isolates from human sources.

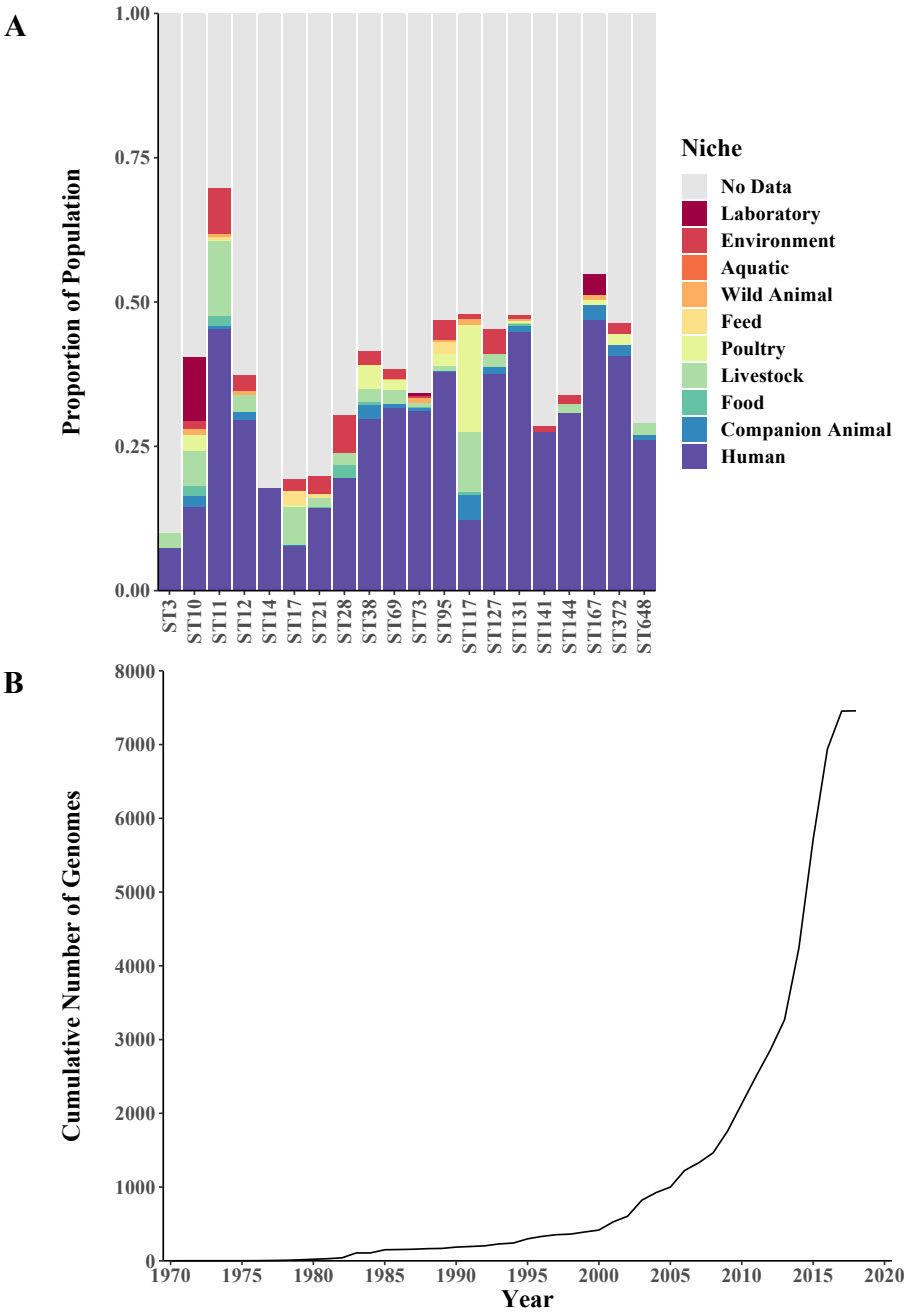

Fig. 3.2 Source niche and time of isolation metrics for the dataset. A) The proportion of the dataset that came from various source niches for each sequence type. Bars are coloured depening on source niche, grey indicates that no data was available. B) Cumulative number of genomes assemblies over time.

Table 3.2 Source Niches with the highest proportion for each Sequence Type

Sequence Type	Highest 3 source niches (descending order)		
ST3	No Data	Human	Livestock
ST10	No Data	Human	Laboratory
ST11	Human	No Data	Livestock
ST12	No Data	Human	Livestock
ST14	No Data	Human	Livestock
ST17	No Data	Human	Livestock
ST21	No Data	Human	Environment
ST28	No Data	Human	Environment
ST38	No Data	Human	Poultry
ST69	No Data	Human	Livestock
ST73	No Data	Human	Wild Animal
ST95	No Data	Human	Environment
ST117	No Data	Poultry	Human
ST127	No Data	Human	Environment
ST131	No Data	Human	Companion Animal
ST141	No Data	Human	Environment
ST144	No Data	Human	Livestock
ST167	Human	No Data	Laboratory
ST372	No Data	Human	Companion Animal
ST648	No Data	Human	Livestock

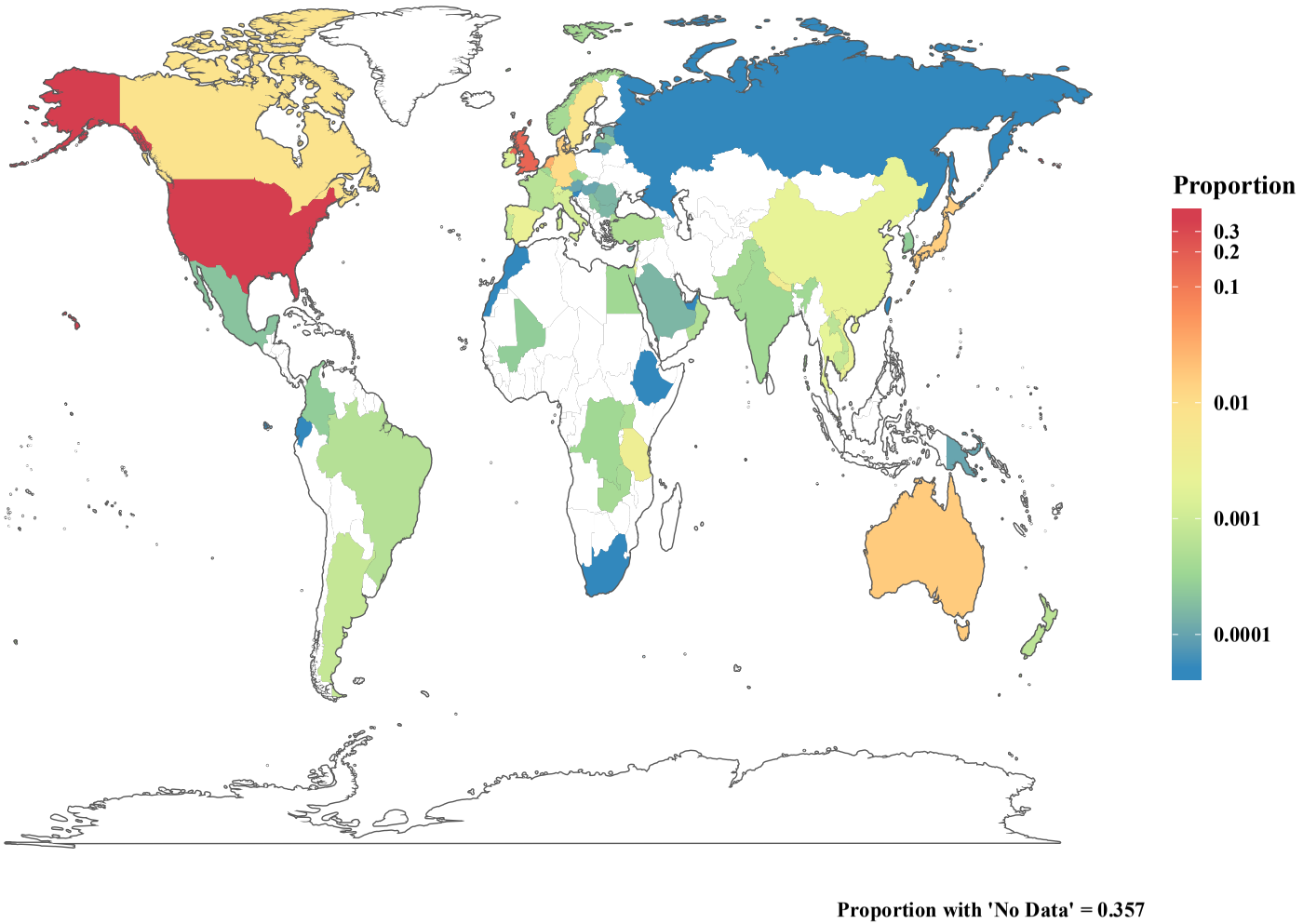

Fig. 3.3 Geographic distribution of the dataset. Countries are coloured based on the their contribution to the dataset . Note that the colour scale is logarithmic.

3.2.2 Population Structure and Antibiotic Resistance

Genomes from the same ST group cluster phylogenetically and major phylogroup divisions are represented in the data

A phylogenetic tree of the entire dataset was produced using the Mashtree script [78] which utilises the Mash program [108] to approximate genome similarity. This approach is not as accurate as a core genome SNP distance tree but is amenable to large datasets. The resulting tree clusters each ST together as expected (Figure 3.4). Moreover the ST groups cluster into their correct phylogroups. The tree also illustrates the highly clonal nature of ST11 with lots of short branches emanating from a central point. Contrasting to the ST10 population which has many longer branches and clear population stratification. ST167 is within the ST10 clonal complex as expected. Moreover the population structure of ST131 is also evident with the canonical clades A, B and C clearly visible. ST131 has also been correctly placed within the larger B2 phylogroup with the other ExPEC populations such as ST73 and ST95.

Antibiotic resistance genes are concentrated in 5 ST groups

The antibiotic resistance gene carriage for every genome in the dataset was assessed using ABRicate with the ResFinder database (March 2018) [168]. The results were summarised as the average number of resistance genes per genome (Figure 3.5a) and the proportion of the population that had 2 or more resistance genes per genome (Figure 3.5b). The data are presented as resistance genes of any type against the number or proportion of β-lactamase genes. Both the average number of genes and proportion of the population carrying multiple resistance genes display a significant positive correlation between resistance genes of all classes and those that are β-lactamases. Indicating that the presence of multiple resistance genes can indicate the presence of β-lactamase genes. This data also highlights the high rates of antibiotic resistance observed in specific populations of *E. coli*. Of note is the ST167 lineage which has an average of 13.9 resistance genes per genome and 2.3 β-lactamase genes

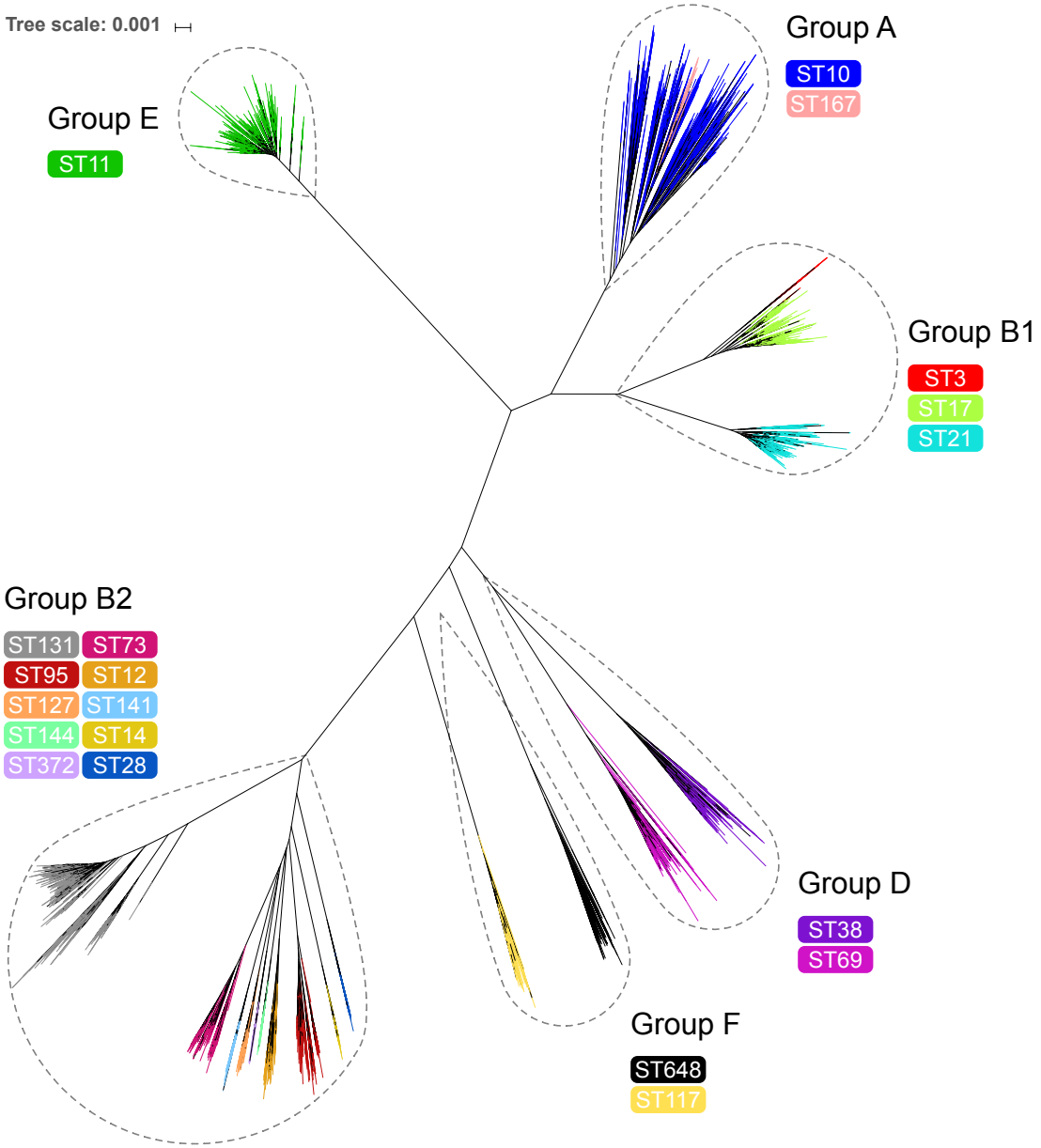

Fig. 3.4 Phylogenetic tree of entire dataset produced using MashTree. Tip branches are coloured according to Sequence Type. Phylogroups are annotaed with dotted bubbles.

per genome. Carriage of resistance genes in the ST167 lineage is also high with 100% of the samples in the dataset containing at least 2 resistance genes and 97% of the samples containing at least 1 β-lactamase gene. At the other end of the spectrum is ST11 with an average of 1.9 resistance genes per genome and an average of 0.07 β-lactamase genes. Only 14% of the ST11 population has 2 or more resistance genes and 6% have at least one β-lactamase gene. It is clear that the majority of the *E. coli* population possesses, on average, less than 5 resistance genes and less than 1 β-lactamase gene. Again the majority of the populations examined have a carriage rate less than 80% for 2 or more resistance genes and less than 60% for 1 or more β-lactamase gene. There are 5 notable lineages with higher rates of resistance gene carriage: ST38, ST69, ST167, ST131 and ST648. These lineages do not share a single phylogroup nor are they clustered on the phylogenetic tree.

Assayed lineages all have open pangenomes with the largest observed pangenome observed in ST10

To elucidate the factors permitting the high rates AMR gene carriage in specific lineages, pangenomes for each lineage were constructed. Roary was used to generate the pangenomes with default parameters [112]. Roary generates files that represent sampling of the pangenome during the analysis process, these files can be analysed to give insight into the pangenome dynamics (Figure 3.6). Examining the rarefaction curves of the pangenomes it is evident that ST10 possesses the largest pangenome with a total of 72,181 genes followed by ST131 and ST11 with 43,422 and 38,569 genes respectively. This is despite ST11 genomes being the most abundant in the dataset; it is therefore apparent that number of genomes used to construct the pangenome does not strictly correlate with pangenome size. All the pangenomes constructed appear to be open pangenomes as supported by power law regression analysis (Table 3.3). None of the pangenomes have an α value greater than 1, indicating that they are all open pangenomes exhibiting "infinite" genetic diversity. The most open pangenome

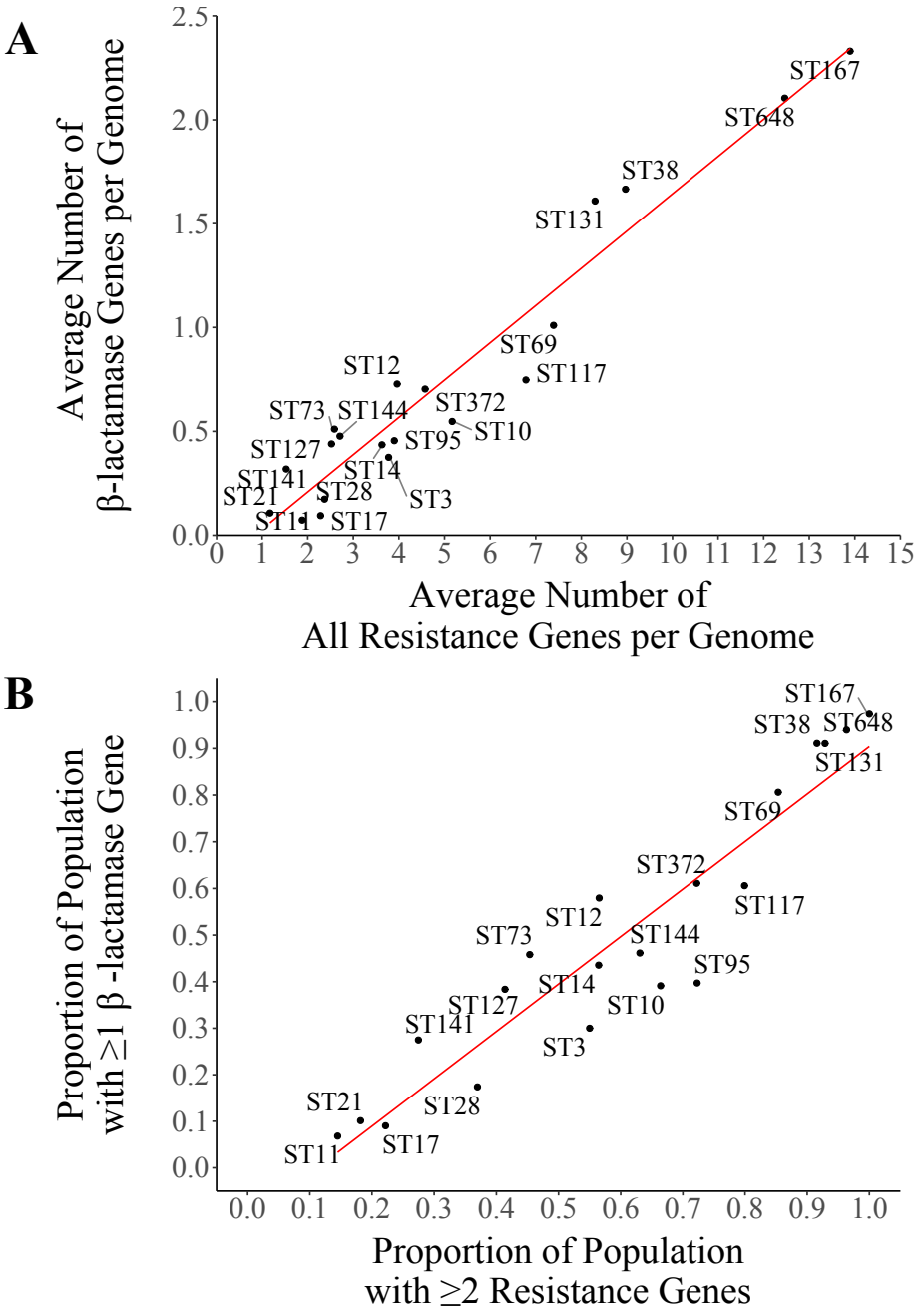

Fig. 3.5 Resistance gene carriage by lineage. A) Average number of β-lactamase genes per genome against the average number of resistance genes of all classes. Red line indicates linear regression line (R^2 value of 0.937, P value of 3.16×10^{-12}) B) The proportion of the population that has one or more β-lactamase gene against the proportion of population that has 2 or more resistance genes per genome. Red line indicates linear regression line (R^2 value of 0.879, P value of 1.07×10^{-9})

is observed in ST10 with an α value of 0.495 whilst the least open pangenome is observed in ST3 with an α value of 0.954. The core genomes can also be visualized using the output from Roary. Their dynamics are the inverse to the total size of the pangenome (Figure 3.7). Specifically ST10 possesses the smallest core genome with 1,761 genes followed by ST131 and ST11 with 2,700 and 3,114 genes respectively. The dynamics of the core genome display an oscillating pattern that repeats every 100 genomes. This is due to Roary classifying a gene as core if it is present in 99% of the samples hence every 100 genomes some genes are re-classified as core. A 99% threshold is used, rather than 100%, to account for potential assembly errors or missing annotations. Plotting the size of the pangenome against the average number of resistance genes per genome reveals that there is no significant correlation (Figure 3.8). Nor is there any correlation between pangenome size and the proportion of the ST populations that carries 2 or more resistance genes.

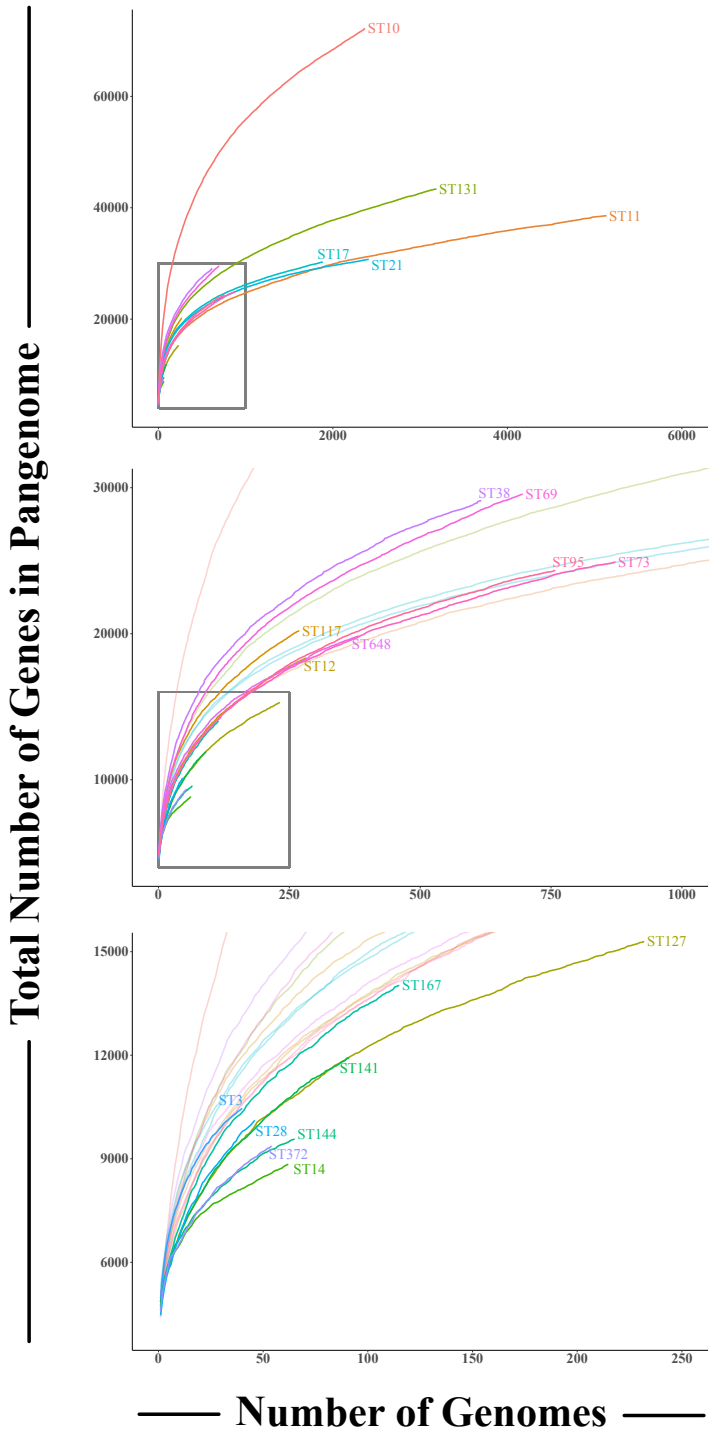

Fig. 3.6 Pangenome rarefaction curves. Average number of genes in the pangenome is plotted against the number of genomes used to construct the pangenome (n=10). Lines are coloured based on sequence type. Boxes indicate zoomed in region visible in subsequent panels. Error bars are omitted for clarity.

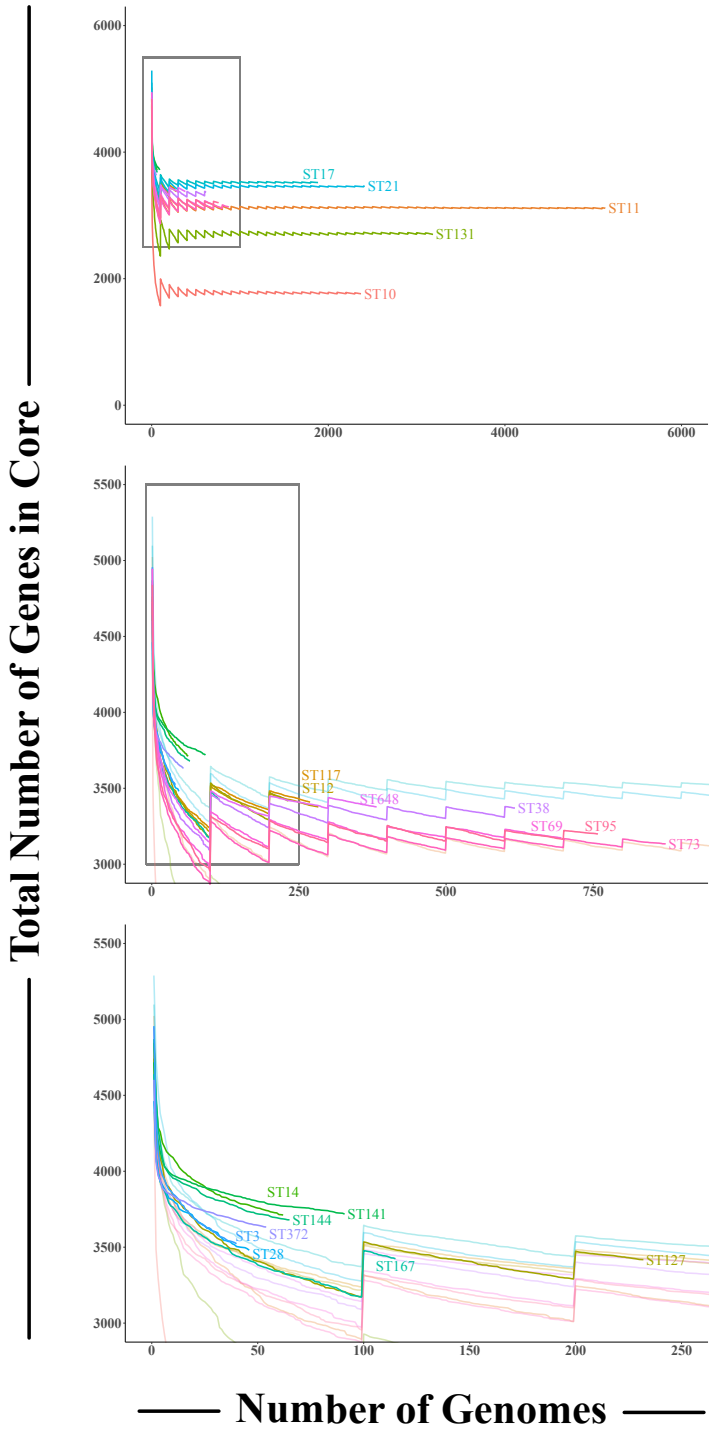

Fig. 3.7 Core pangenome dynamic curves. Average number of genes in the core genome plotted against the number of genomes used to construct the pangenome (n=10). Lines are coloured based on population. Boxes indicate zoomed in region visible in subsequent panel. Error bars are omitted for clarity. Core genes are defined as being present in 99% of samples hence there is an oscillating pattern with every 100th genome.

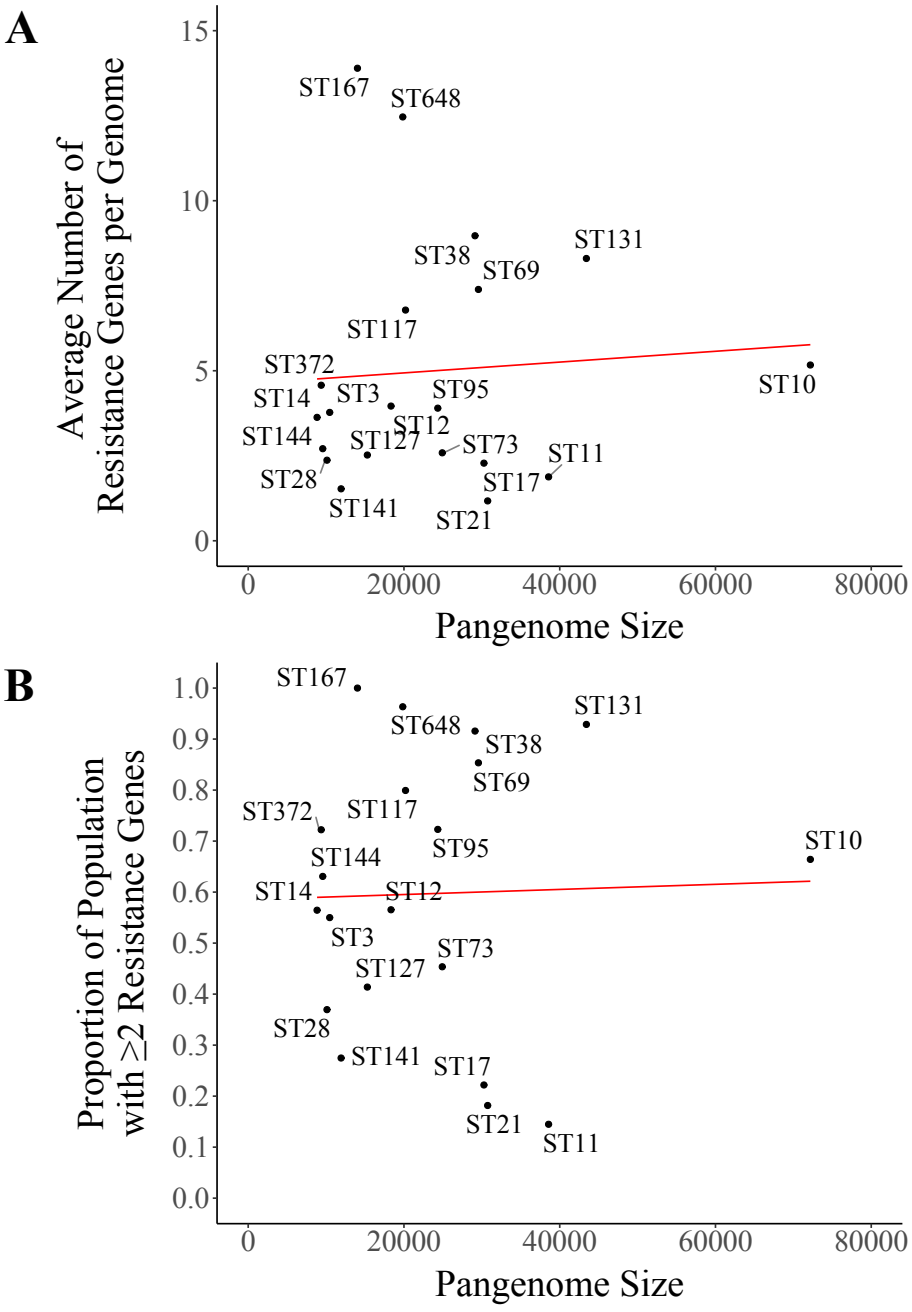

Fig. 3.8 Correlation between pangenome size and carriage of resistance genes. A) Average number of resistance genes per genome against the pangenome size. Red line indicates linear regression line (R^2 value of 0.00462, P value of 0.776). B) The proportion of the population that carries more than one resistance genes against pangenome size. Red line indicates linear regression line (R^2 value of 8.05×10^{-4}, P value of 0.905)

Table 3.3 Pangenome α values. Values greater than 1 indicate a closed pangenome whilst those less than 1 indicate an open pangenome. Populations are ordered by ascending α value.

Sequence Type	Intercept	α - value
ST10	869.3	0.495
ST167	535.6	0.507
ST141	466.8	0.513
ST69	668	0.532
ST38	714.8	0.537
ST127	500.8	0.548
ST117	712.6	0.554
ST12	622.5	0.562
ST648	630.5	0.570
ST73	593.3	0.572
ST131	682.7	0.572
ST95	658.9	0.587
ST11	578.5	0.597
ST21	635.7	0.604
ST17	686.9	0.608
ST28	1114	0.821
ST372	990.4	0.850
ST144	1059.5	0.924
ST14	930.7	0.936
ST3	1541.8	0.954

3.2.3 Functional Pangenomics

The pangenomes presented here were generated using Roary's default parameters, one of which is a 95% identity cut off. Specifically, if the protein sequence of two genes differs by more than 5% then they will be called as separate genes and given separate entries in the pangenome output. As discussed previously this is a highly conservative identity cut-off and will therefore separate alleles of the same gene. The pangenomes were functionally annotated using eggNOG-Mapper [65]. This tool matches query peptide sequences to its database of orthologous groups, which are associated with functional annotations. The associated functional annotation is provided in multiple forms including Clusters of Orthologous Groups (COG) categories and Gene Ontology (GO) terms (Table 3.4).

Table 3.4 COG Categories

COG Category	
A	RNA processing and modification
B	Chromatin structure and dynamics
C	Energy production and conversion
D	Cell cycle control, cell division, chromosome partitioning
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
G	Carbohydrate transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
J	Translation, ribosomal structure and biogenesis
K	Transcription
L	Replication, recombination and repair
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, and chaperones
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport, and catabolism
R	General function prediction only
S	Function unknown
T	Signal transduction mechanisms
U	Intracellular trafficking, secretion, and vesicular transport
V	Defense mechanisms
W	Extracellular structures
Y	Nuclear structure
Z	Cytoskeleton

Hypothetical Proteins are abundant in all lineages

Greater than 50% of each pangenome, with the exception of ST14 and ST372, is composed of hypothetical proteins, highlighting the huge amount of genetic content for which we have no functional information. The proportion of hypothetical proteins displays a positive correlation with the size of the pangenome (Figure 3.9a). This result is mirrored by the functional annotation; the larger the pangenome the smaller the proportion of assigned COG category (Figure 3.9b). Fortunately the proportion of each pangenome assigned a COG category is still greater than 50%.

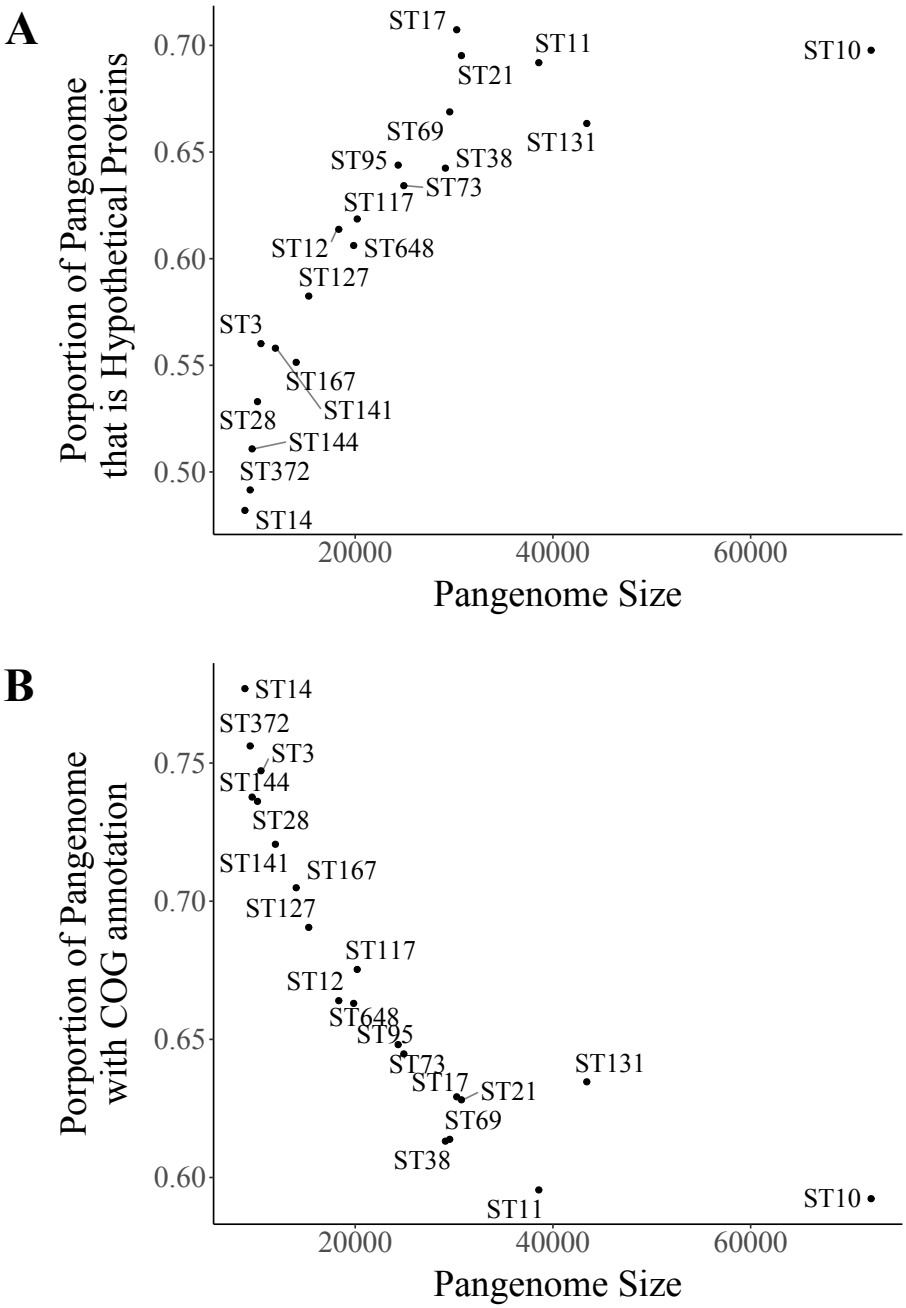

Fig. 3.9 Functional annotation statistics. A) Proportion of the pangenome that is hypothetical proteins against the pangenome size. B) Proportion of the pangenome that is assiged to a COG category against pangenome size.

The COG category 'G – Carbohydrate Transport and Metabolism' displays a significant correlation with antibiotic resistance gene abundance

The COG category with the largest proportion of assigned genes is 'S – Function Unknown' for all the populations (Figure 3.10a). The proportion of the pangenome ranges from 0.278 to 0.441. This is not unsurprising given the high degree of hypothetical proteins present in the pangenome and again reflects the large amount of genetic content for which no biological function has been determined. The second most abundant category in all the pangenomes is 'L – Replication, Recombination and Repair', the proportion ranges from 0.107 to 0.172. The least abundant categories are those specific to eukaryotes, namely 'Y – Nuclear Structure'. The core genome has 'S – Function Unknown' as the most abundant with proportions ranging from 0.208 to 0.236 (Figure 3.10b). The second most abundant category in the core genome is 'G – Carbohydrate Transport and Metabolism', which ranges from 0.0823 to 0.0983. Again, 'S – Function Unknown' is the most abundant category in the accessory pangenome ranging from 0.316 to 0.499 (Figure 3.10c), followed by 'L – Replication, Recombination and Repair' which ranges from 0.183 to 0.245.

In order to determine which biological processes were important for carriage of antibiotic resistance we correlated the proportion of the pangenome annotated with a specific COG category against the proportion of that lineage that possesses multiple resistance genes. To determine which correlations are significant linear regression was performed for all the categories with multiple testing correction. When examining the entire pangenome there was no significant correlation between any COG category and antibiotic resistance (Table 3.5). When examining the core genome there was a significant positive correlation for 'C – Energy Production and Conversion'; as the proportion of the core genome assigned to energy production increases so does the population's carriage of resistance (Table 3.6 and Figure 3.11). Whilst in the accessory genome there is a significant correlation between 'S – Function Unknown' and 'G – Carbohydrate Transport and Metabolism'; the former is negatively cor-

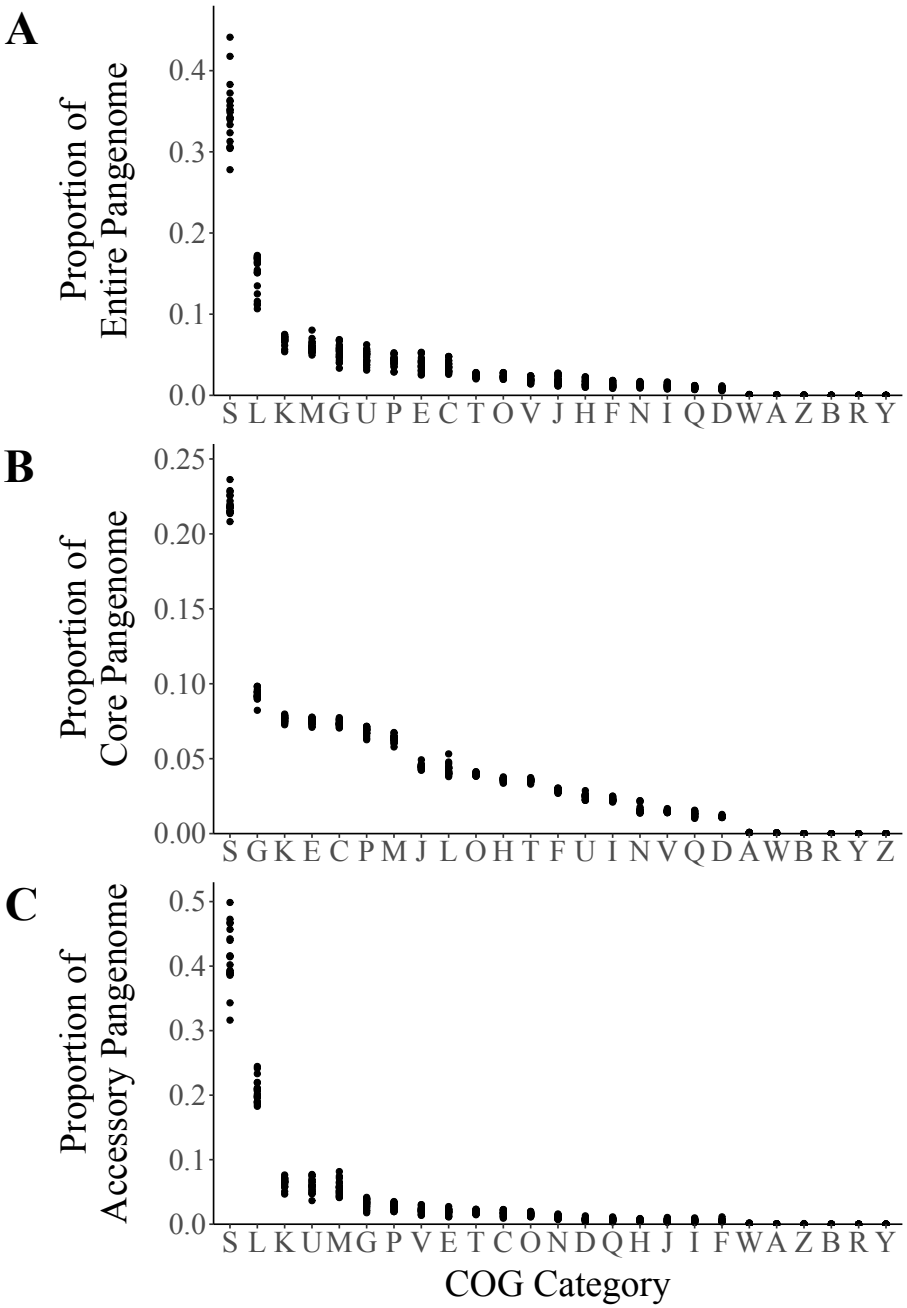

Fig. 3.10 COG category proportions for the different pangenome compartments. Categories are ordered by decreasing average proportion. Dots indicate invidiaul populations. A) Entire pangenome. B) Core pangenome. C) Accessory pangenome.

related while the latter is positively correlated (Table 3.7 and Figure 3.12). The biological implication of the negative correlation between 'S - Unknown Function' and carriage of resistance is unclear, however the positive correlation between 'G - Carbohydrate Transport and Metabolism' and resistance carriage implies that MDR populations display increased variation in carbohydrate metabolic genes. This coupled with the increased abundance of energy production and conversion genes in the core genomes of MDR populations points towards a strong role for metabolism in the evolution of these populations.

Table 3.5 COG Category Correlation with Antibiotic Resistance for the Entire Pangenome. Summary statistics for the correlation between the proportion of the entire functional pangenome assigned to a specific COG category and carriage of antibiotic resistance. Rows are ordered by ascending P value. Rows highlighted in bold face are significant after Bonferoni correction, (0.002 cut off). Correlations were assessed using a linear regression.

COG Category		R^2 value	P value
V	Defense mechanisms	0.344370	0.006529
K	Transcription	0.335797	0.007412
S	Function unknown	0.333626	0.007652
U	Intracellular trafficking, secretion, and vesicular transport	0.254748	0.023230
Q	Secondary metabolites biosynthesis, transport, and catabolism	0.223053	0.035496
Z	Cytoskeleton	0.158947	0.081651
B	Chromatin structure and dynamics	0.132056	0.115278
G	Carbohydrate transport and metabolism	0.094336	0.187755
D	Cell cycle control, cell division, chromosome partitioning	0.082114	0.220616
M	Cell wall/membrane/envelope biogenesis	0.073853	0.246429
A	RNA processing and modification	0.062227	0.288854
N	Cell motility	0.052086	0.333153
L	Replication, recombination and repair	0.031910	0.451132
E	Amino acid transport and metabolism	0.012516	0.638648
P	Inorganic ion transport and metabolism	0.012162	0.643470
J	Translation, ribosomal structure and biogenesis	0.008037	0.707008
F	Nucleotide transport and metabolism	0.006222	0.740969
H	Coenzyme transport and metabolism	0.003944	0.792531
C	Energy production and conversion	0.002565	0.832069
O	Post-translational modification, protein turnover, and chaperones	0.002227	0.843393
I	Lipid transport and metabolism	0.000596	0.918636
T	Signal transduction mechanisms	0.000541	0.922445
W	Extracellular structures	0.000106	0.965615

Table 3.6 COG Category Correlation with Antibiotic Resistance for the Core Pangenome. Summary statistics for the correlation between the proportion of the core functional pangenome assigned to a specific COG category and carriage of antibiotic resistance. Rows are ordered by ascending P value. Rows highlighted in bold face are significant after Bonferoni correction (0.002 cut off). Correlations were assessed using a linear regression.

COG Category		R^2 value	P value
C	**Energy production and conversion**	**0.551970**	**0.000175**
H	Coenzyme transport and metabolism	0.374073	0.004162
S	Function unknown	0.340212	0.006944
E	Amino acid transport and metabolism	0.324556	0.008736
F	Nucleotide transport and metabolism	0.269668	0.018953
L	Replication, recombination and repair	0.266286	0.019852
J	Translation, ribosomal structure and biogenesis	0.195656	0.050833
U	Intracellular trafficking, secretion, and vesicular transport	0.173170	0.068003
A	RNA processing and modification	0.132908	0.114024
D	Cell cycle control, cell division, chromosome partitioning	0.112962	0.147383
K	Transcription	0.112861	0.147576
V	Defense mechanisms	0.100577	0.173058
G	Carbohydrate transport and metabolism	0.085015	0.212282
T	Signal transduction mechanisms	0.083462	0.216699
P	Inorganic ion transport and metabolism	0.077312	0.235228
I	Lipid transport and metabolism	0.037179	0.415372
Q	Secondary metabolites biosynthesis, transport, and catabolism	0.034695	0.431698
N	Cell motility	0.024190	0.512611
M	Cell wall/membrane/envelope biogenesis	0.020095	0.551069
O	Post-translational modification, protein turnover, and chaperones	0.005286	0.760666
W	Extracellular structures	0.002217	0.843722

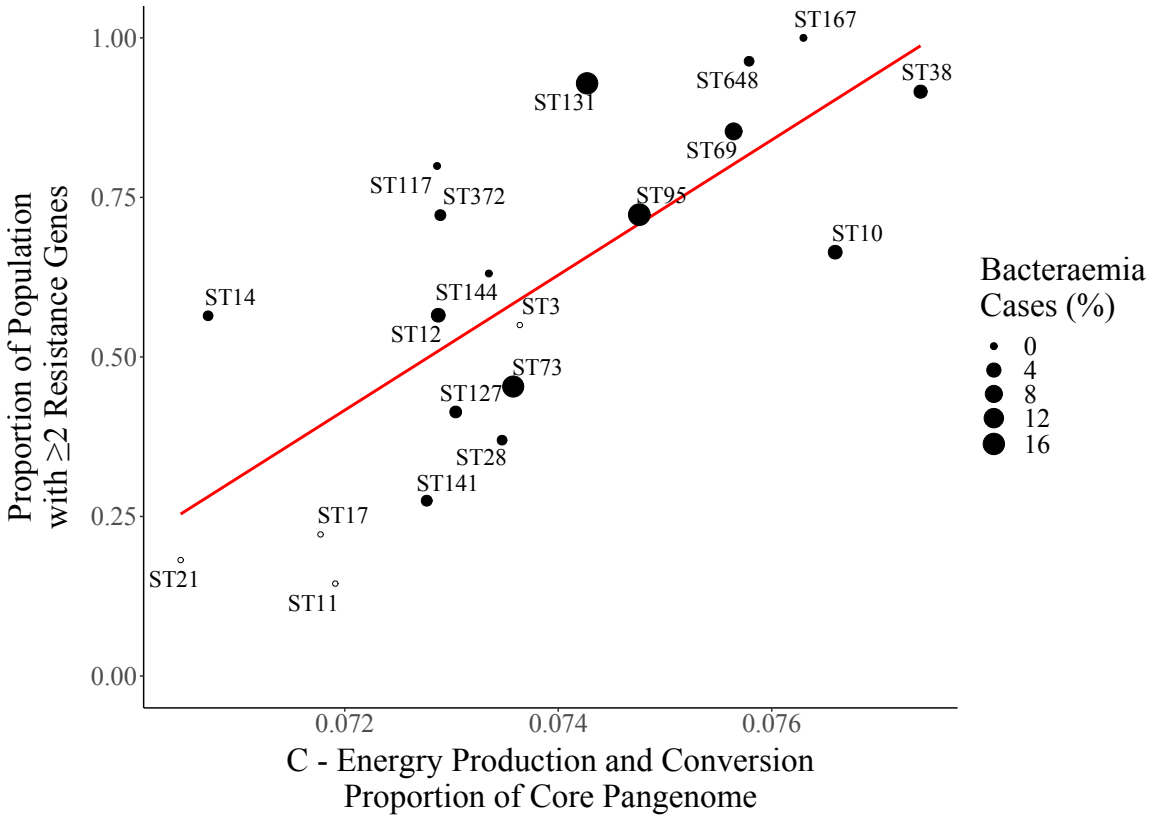

Fig. 3.11 Correlation bewteen proportion of core pangenome annotated with 'Energy Production and Conversion' and carriage of antibiotic resistance. Size of point indicates percentage of bacteraemia cases where that population was isolated, while open points indicate that population was not present in bactaraemia surveillance dataset (BSAC). Red line indicates linear regression (R^2 value of 0.552, P value of 1.75×10^{-3})

Table 3.7 COG Category Correlation with Antibiotic Resistance for the Accessory Pangenome. Summary statistics for the correlation between accessory pangenome proportion of COG category and carriage of antibiotic resistance. Rows are ordered by ascending P value. Rows highlighted in bold face are significant after Bonferoni correction (0.002 cut off).

COG Category		R^2 value	P value
S	**Function unknown**	**0.571741**	**0.000115**
G	**Carbohydrate transport and metabolism**	**0.566855**	**0.000128**
K	Transcription	0.352814	0.005755
V	Defense mechanisms	0.266439	0.019811
U	Intracellular trafficking, secretion, and vesicular transport	0.236117	0.029842
J	Translation, ribosomal structure and biogenesis	0.205106	0.044940
Z	Cytoskeleton	0.176614	0.065049
E	Amino acid transport and metabolism	0.163844	0.076673
H	Coenzyme transport and metabolism	0.148950	0.092823
N	Cell motility	0.143562	0.099463
Q	Secondary metabolites biosynthesis, transport, and catabolism	0.135430	0.110395
B	Chromatin structure and dynamics	0.128920	0.120013
D	Cell cycle control, cell division, chromosome partitioning	0.108560	0.156016
M	Cell wall/membrane/envelope biogenesis	0.104272	0.164937
A	RNA processing and modification	0.089933	0.198934
L	Replication, recombination and repair	0.069529	0.261299
P	Inorganic ion transport and metabolism	0.063713	0.282978
T	Signal transduction mechanisms	0.057549	0.308331
I	Lipid transport and metabolism	0.055378	0.317909
C	Energy production and conversion	0.048072	0.353009
W	Extracellular structures	0.008999	0.690759
O	Post-translational modification, protein turnover, and chaperones	0.003471	0.805119
F	Nucleotide transport and metabolism	0.000266	0.945625

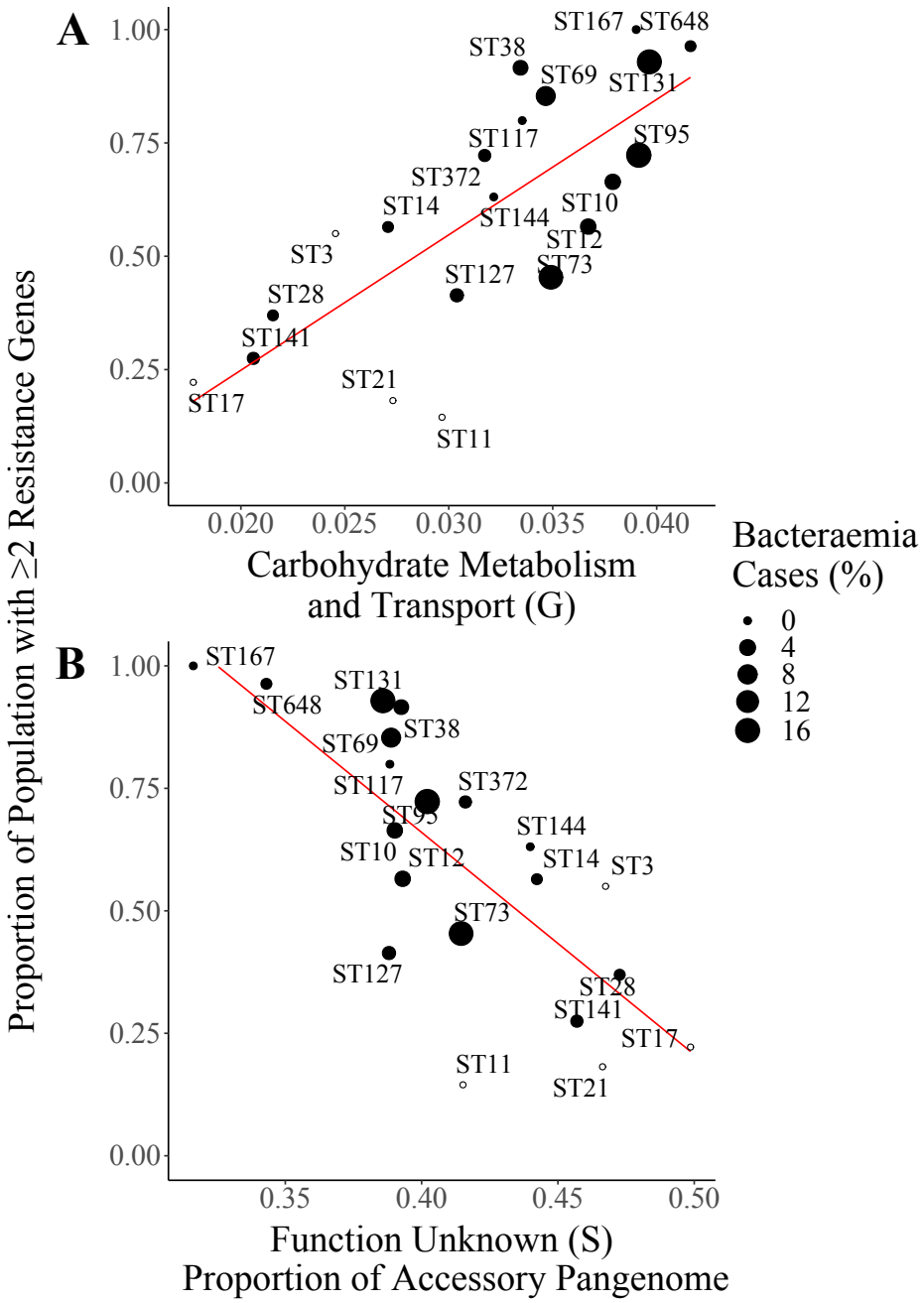

Fig. 3.12 Correlation between proportion of accessory pangenome annotaed with COG category and cariage of antibiotic resistance. Size of point indicates percentage of bacteraemia cases where that population was isolated, while open points indicate that population was not present in bactaraemia surveillance dataset (BSAC). A) Correlation bewteen proportion of core pangenome annoated with 'Carbohydrate transport and metabolism' and carriage of antibiotic resistance. Red line indicates linear regression (R^2 value of 0.567, P value of 1.28×10^{-4}) B) Correlation bewteen proportion of accessory pangenome annoated with 'Function unknown' and carriage of antibiotic resistance. Red line indicates linear regression (R^2 value of 0.572, P value of 1.15×10^{-4})

Metabolic GO Terms are enriched in MDR lineages

Whilst the COG categories are useful for examining the broad processes that are undergoing selection in MDR populations they provide little specific information about biological processes and pathways. To address this GO term enrichment analysis was employed. The accessory and core genomes of MDR lineages (ST38, ST69, ST131, ST167 and ST648) were compared to their counterparts from the rest of the genome dataset. The analysis determined which GO terms appeared significantly more frequently in the MDR pangenomes compared to the non-MDR pangenomes. The most significantly enriched GO term in the core pangenome was 'Defence response to virus' (Table 3.8). There are multiple and varied catabolic processes that also appear enriched in the core genome. This is consistent with an increase in the COG category for 'Energy Production and Conversion'. Terms referring to phenylpropionate metabolism and transport also occur frequently, alongside terms for the catabolism of amino acids. The most enriched term in the accessory pangenome is 'D-galactonate catabolic' process (Table 3.9). There are several amino acid biosynthetic or catabolic terms also appearing. There are also several terms referring to DNA recombination, transposition and repair. The term β-lactam catabolic process also appears enriched in MDR accessory pangenomes.

Table 3.8 Gene Ontology (GO) terms enriched in the core pangenomes of multi-drug resistant lineages.

	GO ID	Term	Annotated	Significant	Expected	P value (Fisher)
1	GO:0051607	defense response to virus	43	22	10.82	0.00022
2	GO:0019380	3-phenylpropionate catabolic process	46	18	11.58	0.02575
3	GO:0019622	3-(3-hydroxy)phenylpropionate catabolic process	22	10	5.54	0.03083
4	GO:0046271	phenylpropanoid catabolic process	22	10	5.54	0.03083
5	GO:0010124	phenylacetate catabolic process	98	33	24.67	0.03692
6	GO:0009440	cyanate catabolic process	29	12	7.30	0.04092
7	GO:0043571	maintenance of CRISPR repeat elements	17	8	4.28	0.04175
8	GO:0009313	oligosaccharide catabolic process	168	45	42.29	0.05025
9	GO:0043488	regulation of mRNA stability	30	12	7.55	0.05302
10	GO:1901606	alpha-amino acid catabolic process	657	170	165.37	0.05896
11	GO:0046657	folic acid catabolic process	18	8	4.53	0.05899
12	GO:0006548	histidine catabolic process	2	2	0.50	0.06335
13	GO:0032781	positive regulation of ATPase activity	10	5	2.52	0.08011
14	GO:0009447	putrescine catabolic process	115	36	28.95	0.08153
15	GO:0006115	ethanol biosynthetic process	33	12	8.31	0.10282
16	GO:0016579	protein deubiquitination	8	4	2.01	0.11617
17	GO:0015814	p-aminobenzoyl-glutamate transport	8	4	2.01	0.11617
18	GO:0042920	3-hydroxyphenylpropionic acid transport	11	5	2.77	0.11736
19	GO:0015731	3-hydroxyphenyl propanoate transport	11	5	2.77	0.11736
20	GO:0015765	methylgalactoside transport	18	7	4.53	0.14289
21	GO:0015769	melibiose transport	18	7	4.53	0.14289
22	GO:0015682	ferric iron transport	55	16	13.84	0.15212
23	GO:0042942	D-serine transport	60	19	15.10	0.15583
24	GO:0019509	L-methionine salvage from methylthioadenosine	3	2	0.76	0.15817
25	GO:0015700	arsenite transport	9	4	2.27	0.16891

Table 3.9 Gene Ontology (GO) terms enriched in the accessory pangenomes of multi-drug resistant lineages.

	GO ID	Term	Annotated	Significant	Expected	P value (Fisher)
1	GO:0034194	D-galactonate catabolic process	150	88	47.46	7.8e-12
2	GO:0000162	tryptophan biosynthetic process	82	56	25.94	1.0e-11
3	GO:0008652	cellular amino acid biosynthetic process	1166	483	368.89	1.9e-10
4	GO:0019270	aerobactin biosynthetic process	246	125	77.83	3.0e-10
5	GO:0006310	DNA recombination	4216	1451	1333.84	2.0e-05
6	GO:0030655	beta-lactam antibiotic catabolic process	193	88	61.06	3.3e-05
7	GO:0006281	DNA repair	1893	644	598.90	3.3e-05
8	GO:0015940	pantothenate biosynthetic process	43	29	13.60	6.2e-05
9	GO:0006523	alanine biosynthetic process	29	20	9.17	6.3e-05
10	GO:0019540	siderophore biosynthetic process from catechol	107	50	33.85	0.00011
11	GO:0055072	iron ion homeostasis	340	128	107.57	0.00011
12	GO:0006313	transposition, DNA-mediated	3169	1095	1002.60	0.00014
13	GO:0000920	septum digestion after cytokinesis	42	25	13.29	0.00018
14	GO:0006798	polyphosphate catabolic process	19	14	6.01	0.00020
15	GO:0006417	regulation of translation	165	55	52.20	0.00021
16	GO:0006261	DNA-dependent DNA replication	678	252	214.50	0.00037
17	GO:0009651	response to salt stress	47	24	14.87	0.00057
18	GO:0032506	cytokinetic process	62	34	19.62	0.00064
19	GO:0010350	cellular response to magnesium starvation	92	44	29.11	0.00085
20	GO:0000027	ribosomal large subunit assembly	33	19	10.44	0.00183
21	GO:0006565	L-serine catabolic process	52	27	16.45	0.00186
22	GO:0006531	aspartate metabolic process	63	28	19.93	0.00200
23	GO:0042177	negative regulation of protein catabolic process	98	45	31.00	0.00213
24	GO:0006434	seryl-tRNA aminoacylation	10	8	3.16	0.00234
25	GO:0015820	leucine transport	12	9	3.80	0.00257

3.2.4 Sources of Genetic Variation

Paralogs do not contribute to the significant association between metabolism and antibiotic resistance

Roary was used to construct the pangenomes using default settings, one of which is to split paralogs in the output. If the programme identifies multiple copies of the same gene from the genome it will attempt to split these copies into separate entries in the pangenome. To achieve this Roary examines the gene neighbourhood (the 5 genes upstream and downstream of a gene) to determine which pangenome cluster it should be placed into. By disabling this process it is possible to assess the contribution that paralogs have to the pangenome and additionally if they play a role in the previously observed associations. When paralog splitting is disabled there is a decrease in the total pangenome size for all populations assayed (Figure 3.13). The greatest decrease in size, as a proportion of pangenome size, is seen in ST21 (-0.524) followed by ST17 (-0.459) then ST11 (-0.455). All of these populations are EHEC lineages. The greatest increase in core size, as a proportion, is seen in ST10 (0.763) followed by ST131 (0.349) then ST11 (0.281).

To assess whether paralogs, or the paralog splitting functionality of Roary, influenced the observed correlation between antibiotic resistance carriage and specific COG categories the analysis was repeated using paralog-free pangenomes. The previously observed significant positive correlation between 'Energy Production and Conversion' in the core pangenome remains significant despite an increase in the p value (0.000175 vs. 0.00131) (Figure 3.14). The same is true for the correlation between 'Carbohydrate Transport and Metabolism' in the accessory pangenome and antibiotic resistance carriage (p value of 0.0000128 vs. 0.001263) (Figure 3.15a). The correlation between resistance carriage and 'Unknown Function' in the accessory pangenome loses significance (p value of 0.0000115 vs. 0.1353) (Figure 3.15b). This is likely driven by the decrease in ST11, ST21 and ST17 suggesting that the paralogs identified in these populations were of unknown function.

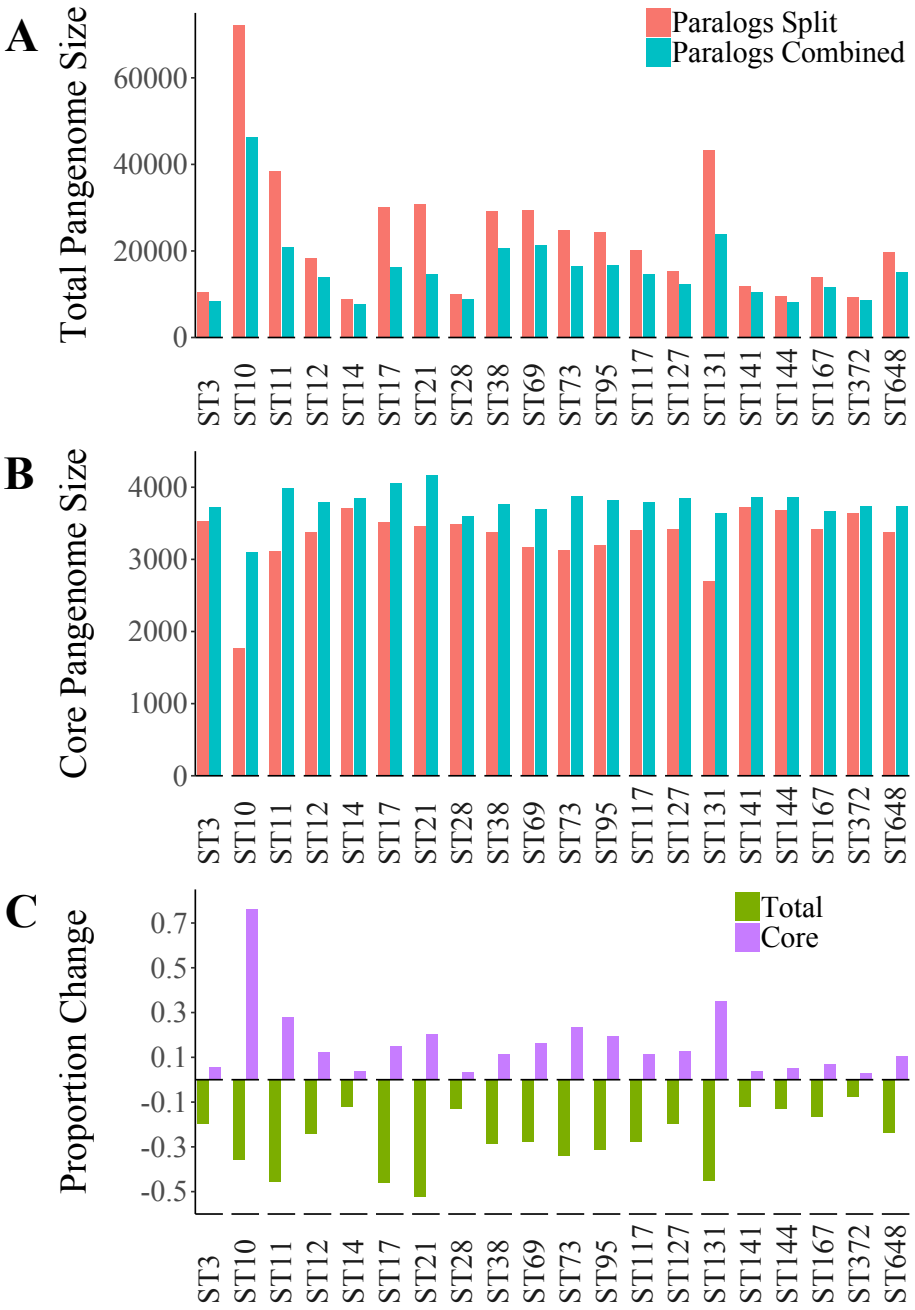

Fig. 3.13 Contribution of paralogs to pangenome size. Bars in red indicate pangenome size when paralos are split by Roary, bars in blue are pangenome size when paralogs are not split by Roary (A & B). A) Total pangenome size for each sequence type. B) Core pangenome size for each sequence type. C) Proportion change in pangenome size for each sequence type. Bars in green indicate change for the total pangenome size while bars in purple indicate change for core pangenome size.

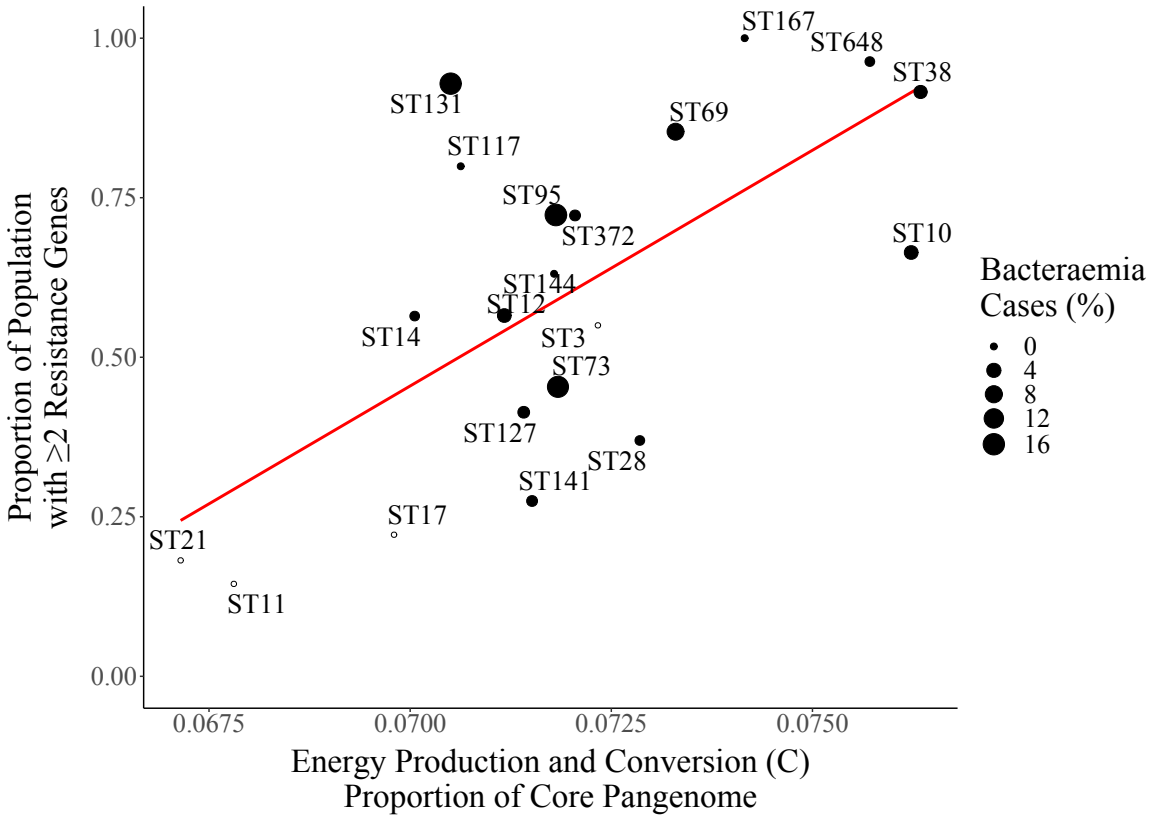

Fig. 3.14 Core Pangenome Correlation with C COG Category when Paralog Splitting is Disabled. Red line represents linear regression line (R^2 value of 0.445, P value of 0.00131). Size of point indicates percentage of bactaraemia cases where that population was isolated, while open points indicate that population was not present in bactaraemia surveillance dataset (BSAC).

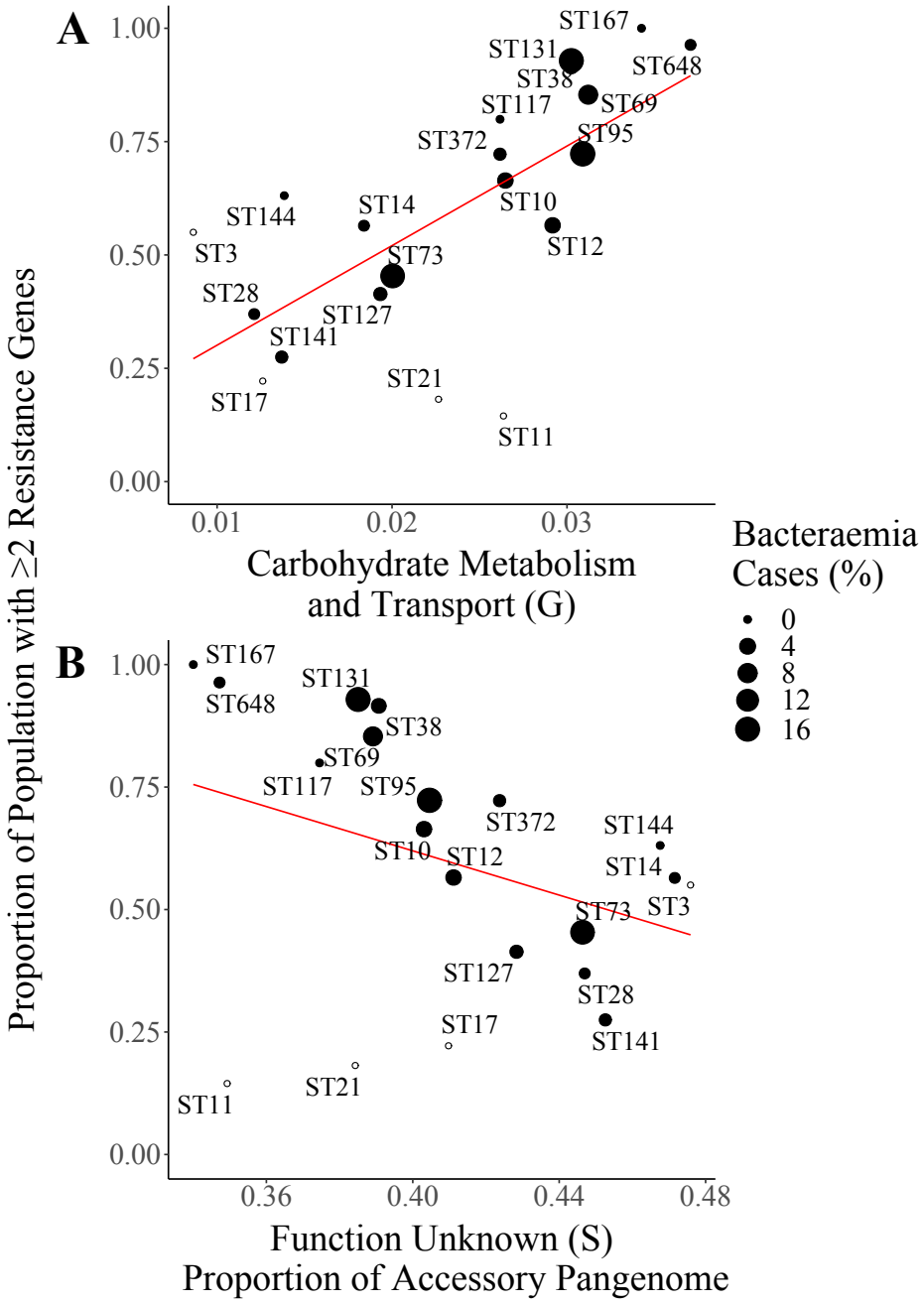

Fig. 3.15 Accessory Pangenome Correlation with COG Category Proportion when Paralog Splitting is Disabled. Size of point indicates percentage of bactaraemia cases where that population was isolated, while open points indicate that population was not present in bactaraemia surveillance dataset (BSAC). A) Correlation with carbohydrate metabolism and tranport. Red line represents linear regression line (R^2 value of 0.4166, P value of 0.001263). B) Correlation with function unknown category. Red line represents linear regression line (R^2 value of 0.07069, P value of 0.1353).

MDR lineages do not show increased diversity in intergenic regions

Another key consideration when examining microbial evolution are Intergenic Regions (IGRs). These pieces of DNA located in the space between genes contain important regulatory elements such as promoters, terminators and non-coding RNAs. These elements can influence the bacterial phenotype, for example it has been demonstrated that bacteria can adapt to the presence of plasmids through mutations in intergenic regions [61]. Moreover the ST131 lineage displays variation in intergenic regions which is strongly linked to CTX-M gene carriage, with different CTX-M types associated with different IGRs [96]. The analysis presented so far has omitted any contribution that these components may make towards the carriage of antibiotic resistance. To address this the programme Piggy was utilised [148]. This programme is capable of identifying IGRs from annotated genome assemblies before performing a clustering analysis, similar to the pangenome pipeline, to identify unique and shared IGRs. Analysing the total number of unique IGR identified with Piggy reveals that there is a strong positive correlation between the size of the pangenome and the number of unique IGRs identified (Figure 3.16a). This is not unexpected as larger pangenomes demonstrate increased diversity, this analysis confirms that this diversity is not limited to coding regions. The relationship between pangenome size and IGR diversity is linear but appears to deviate from a 1 to 1 ratio. There is not a significant correlation between the total number of IGR and carriage of antibiotic resistance (Figure 3.16).

Bacteriophage do not contribute towards the significant association between metabolism and AMR gene prevalence

Bacteriophage are important drivers of variation in bacteria; they are capable of transferring portions of DNA between bacteria via a process called transduction. Phage elements were identified in the pangenome by BLAST searching the translated peptide sequences against a database of phage sequences (PHAST) [170]. Across the entire dataset ST11 had the high-

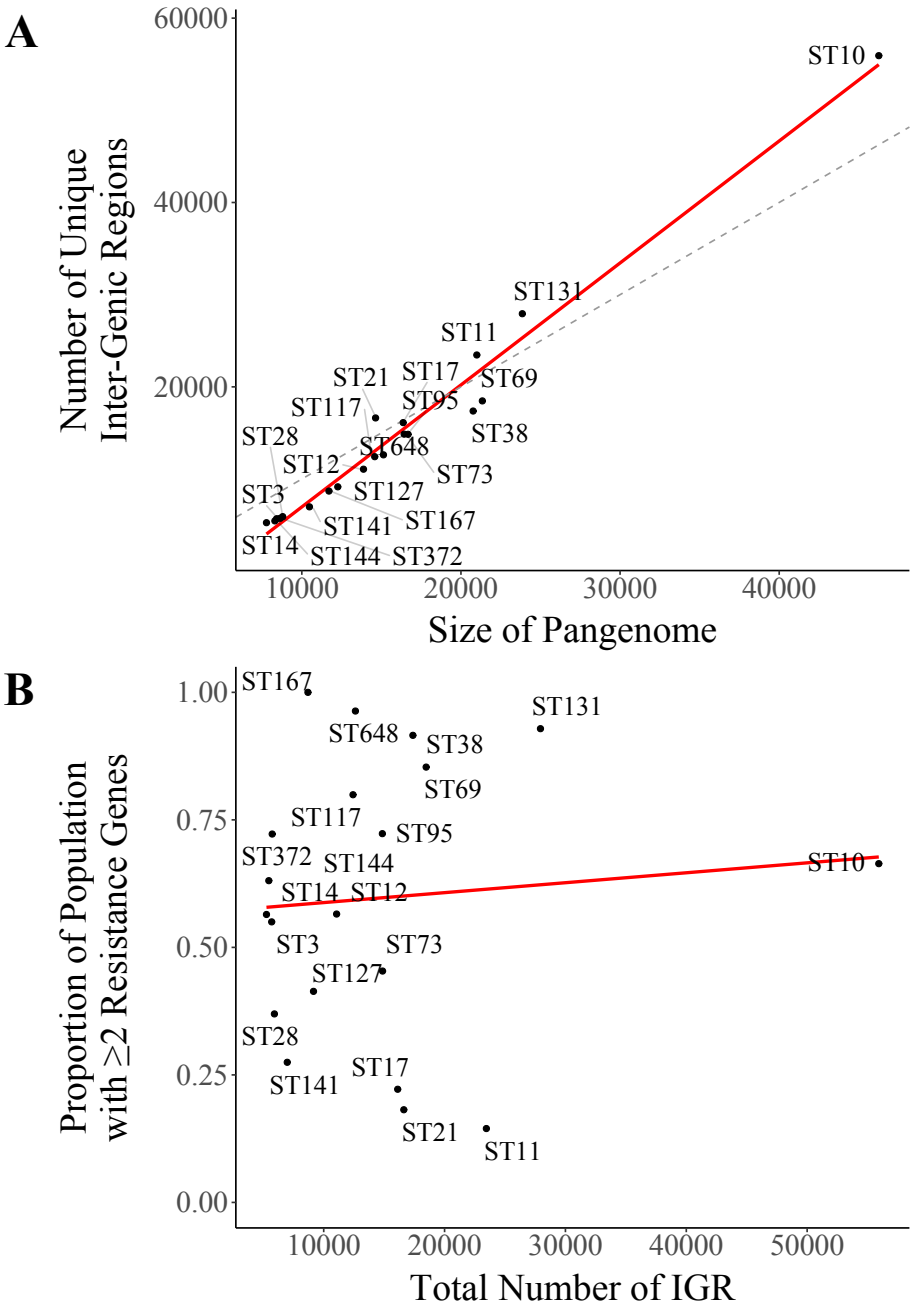

Fig. 3.16 A) Number of unique intergenic regions plotted against pangenome size. Red line indicates linar regression (R^2 value of 0.976, P value of 4.44×10^{-16}). Dotted line follows $x = y$. B) Proportion of population that is anitbiotic resistant against the total number of unique IGR. Red line indicates linar regression (R^2 value of 0.00697, P value of 0.726)

est phage proportion with 0.317 of the pangenome being phage associated (Figure 3.17a). The next highest were ST17 and ST21 with 0.293 and 0.291 respectively. Again, these are EHEC lineages which also displayed similar levels of paralogs when compared to the other populations. For all the lineages assayed the majority of phage were in the accessory genome with fewer elements found in the intermediate pangenome and fewer still in the core genome. The proportion of the pangenome associated with phage elements is significantly negatively correlated with the carriage of antibiotic resistance (Figure 3.17b). This correlation is strongly influenced by the high rates of phage in ST11, ST17 and ST21 combined with their low carriage of antibiotic resistance genes. Highly MDR lineages have comparable proportion of phage elements to the other lineages analysed. To test if phage had any impact on the previously observed correlations between COG categories and carriage of antibiotic resistance genes the analysis was repeated after removing phage associated elements from the pangenomes (Figure 3.18). The positive correlation between proportion of the accessory genome annotated with 'Carbohydrate Transport and Metabolism' and the proportion of the population with more than one resistance gene remains significant despite an increase in the P value (0.000128 with phage elements vs. 0.00568 without phage) (Figure 3.18a). The same is true for the correlation between 'Unknown Function' and resistance carriage (0.000115 with phage versus 0.00955 without phage) (Figure 3.18b).

Levels of recombination are positively correlated with antibiotic resistance gene carriage

Transduction via phage is not the only mechanism that allows bacteria to acquire genetic information. There are multiple mechanisms that permit sharing of DNA amongst bacteria. It is possible to measure some of these mechanisms by looking for signatures of recombination in the genome. For this analysis reference strains for each ST were selected randomly, except for ST131 where a clade C genome was chosen randomly. Genomes were then mapped

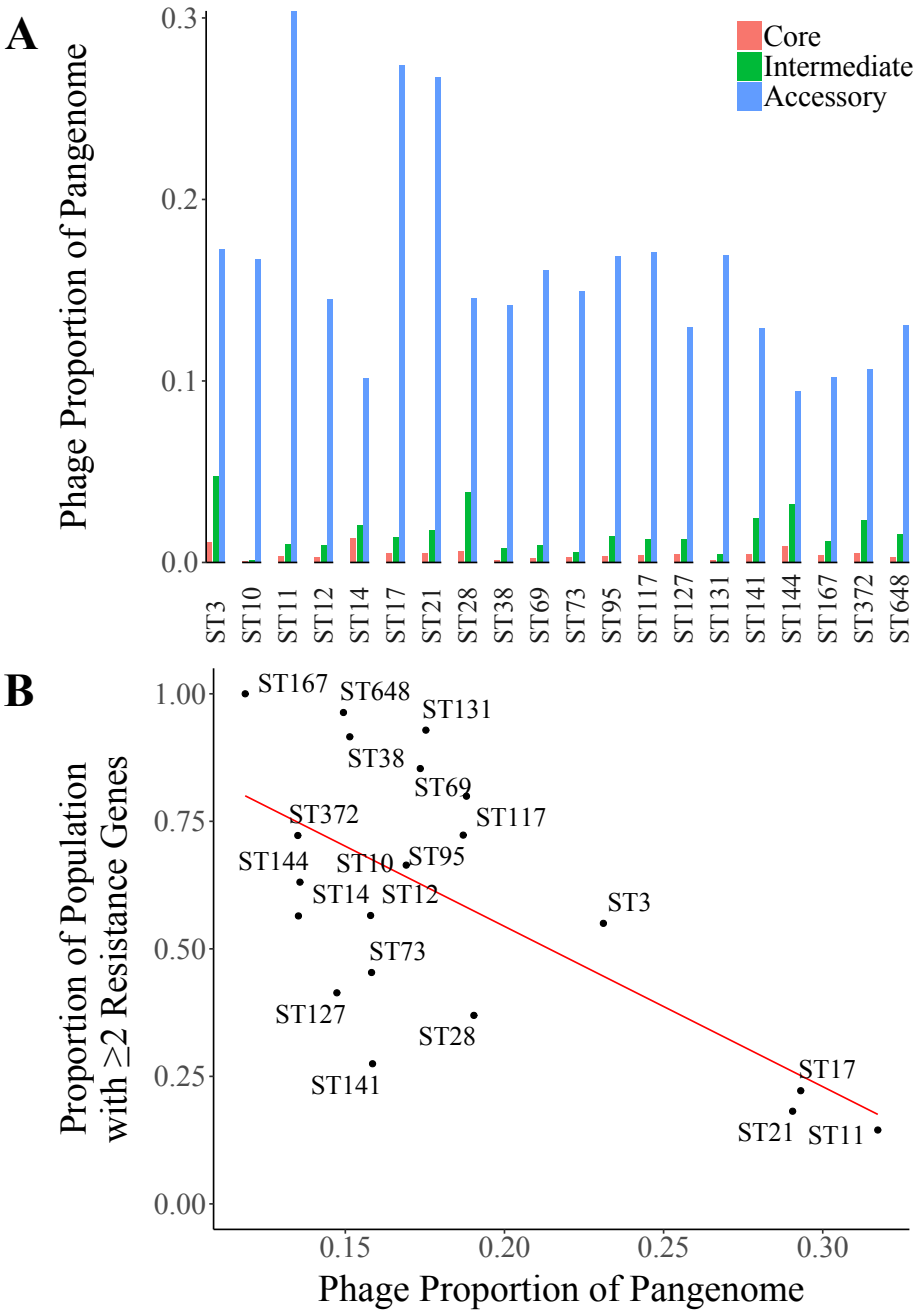

Fig. 3.17 Proportion of pangenome that is associated with phage elements. A) Proportion of pangenome that is associated with phage elements for each sequence type. Bars are coloured depending on pangenome fraction; red indicates core, green indicates intermediate while blue indicates accessory. B) Correlation between proportion of pangenome that is phage associated and carriage of antibiotic resistance. Red line represents linear regression line (R^2 value of 0.433, P value of 0.00161).

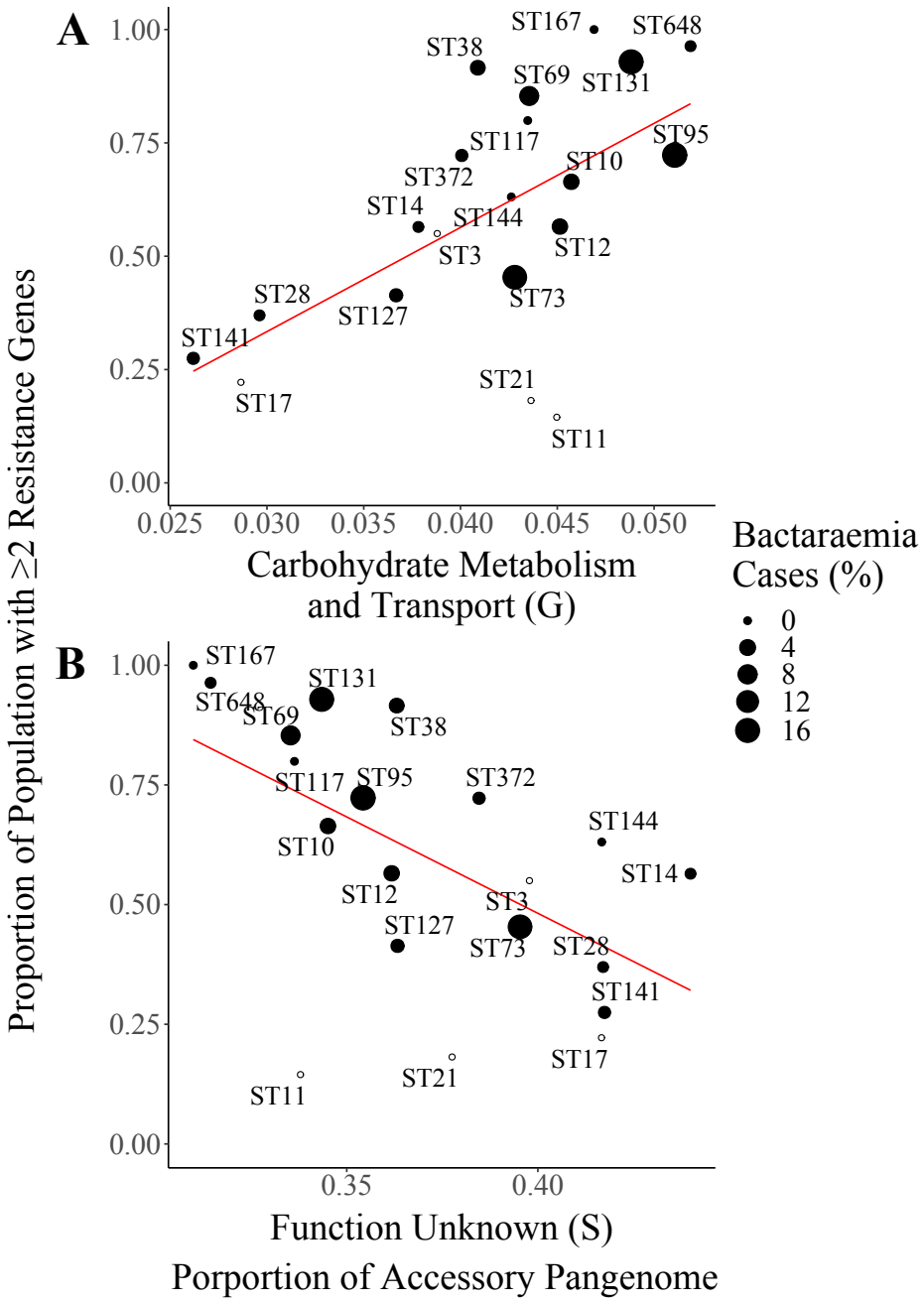

Fig. 3.18 COG Category Correlations After Removing Phage Elements. Size of point indicates percentage of bacteraemia cases where that population was isolated, while open points indicate that population was not present in the bacteraemia dataset (BSAC). A) Correlation with carbohydrate metabolism and tranport. Red line represents linear regression line (R^2 value of 0.0.354, P value of 0.00568). B) Correlation with 'Function Unknown' category. Red line represents linear regression line (R^2 value of 0.0.318, P value of 0.00955).

to the selected reference using SKA [60]. The resulting alignments are then analysed using Gubbins to measure recombination [31]. The ST11 population was excluded as it was not computationally feasible to analyse.

The proportion of the genome in a predicted recombination block was compared between lineages (Figure 3.19a). The lineage with the highest proportion was ST10 with a median of 0.352 followed by ST648 (0.195) then ST117 (0.164). The lineages with the lowest levels of recombination were ST21 with 0.0405 followed by ST144 (0.0488) then ST17 (0.0512). The ST10 lineage was the only one with a median higher than 0.2. This exception is not mirrored by the other phylogroup A lineage: ST167.

The proportion of the genome within a predicted recombination block is significantly positively correlated with resistance gene carriage (Figure 3.19b). A higher proportion of recombination blocks is observed in lineages with higher levels of resistance gene carriage. Notably, the ST10 population does not fit this trend exhibiting a high level of recombination but intermediate level of resistance carriage. By comparing the recombination predictions from Gubbins with the genome annotation of the randomly selected reference strains it is possible to manually examine which genes are in, or adjacent to, the blocks or recombination. Whilst this is possible for all the linegaes, only ST73 and ST131 were examined. These populations were chosen as ST131 is a highly MDR lineage that is globally disseminated whilst ST73 shares several properties with ST131 (phylogroup, pathogenicity) it is rarely found to be MDR. The differing levels of recombination between the 2 populations are evident with numerous blocks predicted in ST131 whilst there are fewer in ST73 (Figure 3.20). The two lineages share functional genes associated with recombination blocks such as those involved in antibiotic resistance, fimbria and capsule genes. Genes associated with phage appear multiple times in proximity or within recombination blocks in both populations. Genes involved in amino acid metabolism also appear in both populations whilst genes involved in oligosaccharide catabolism are only evident in ST131.

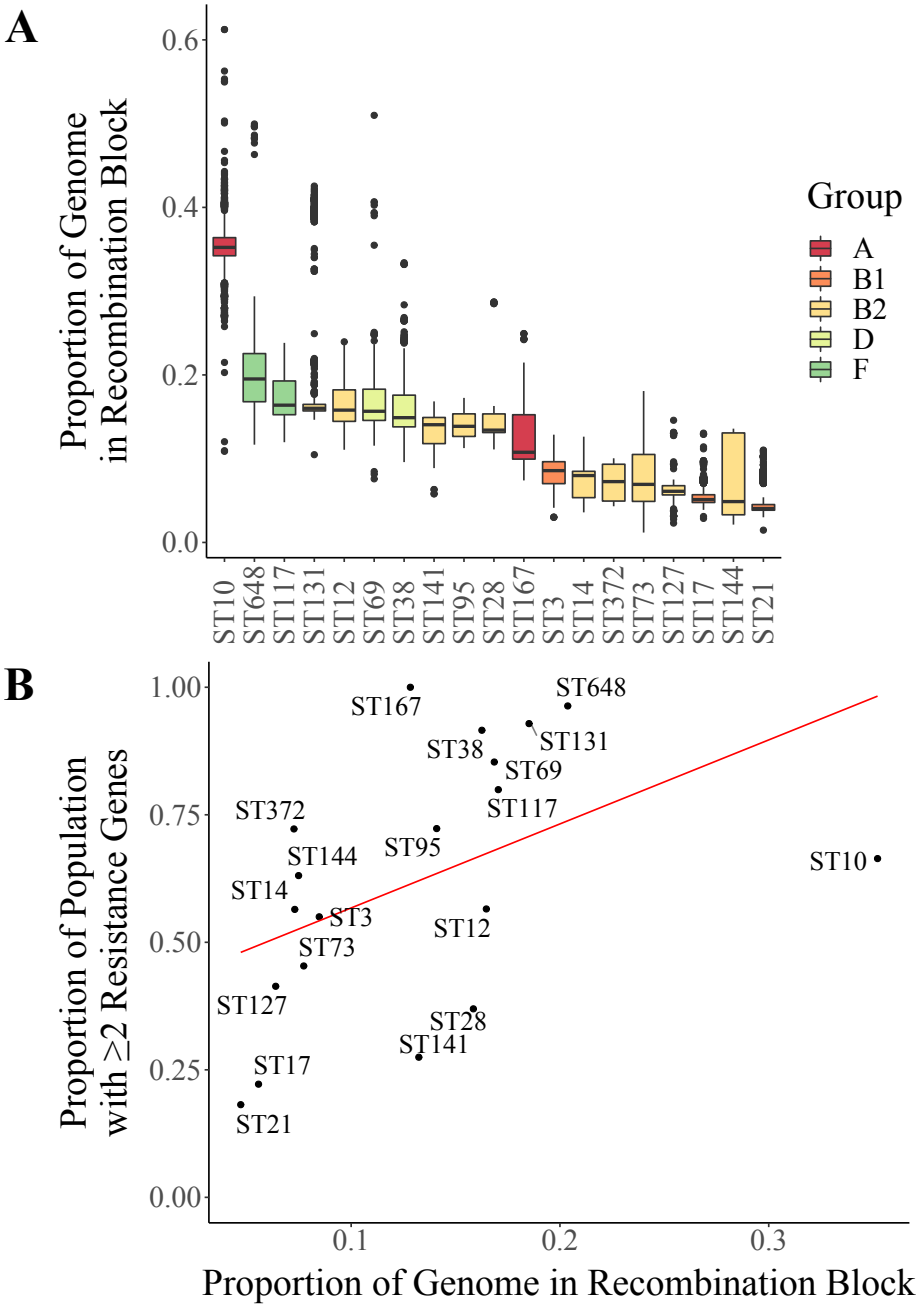

Fig. 3.19 Percentage of genome in a recombination block. A) Percentage of genome in a recombination block for each sequence type. Bold central line indicates median for population. The lower and upper bounds for the box indicate the 25th and 75th percentile. Lines indicate values within 1.5x the interquartile range whilst dots indicate values outside 1.5x the interquartile range. Boxes are coloured based upon the phylogroup. B) Proportion of population with 2 or more resistance genes against the average proportion of the genome in a recombination block. Red line indicates linear regression line (R^2 value of 0.221, P value of 0.0422).

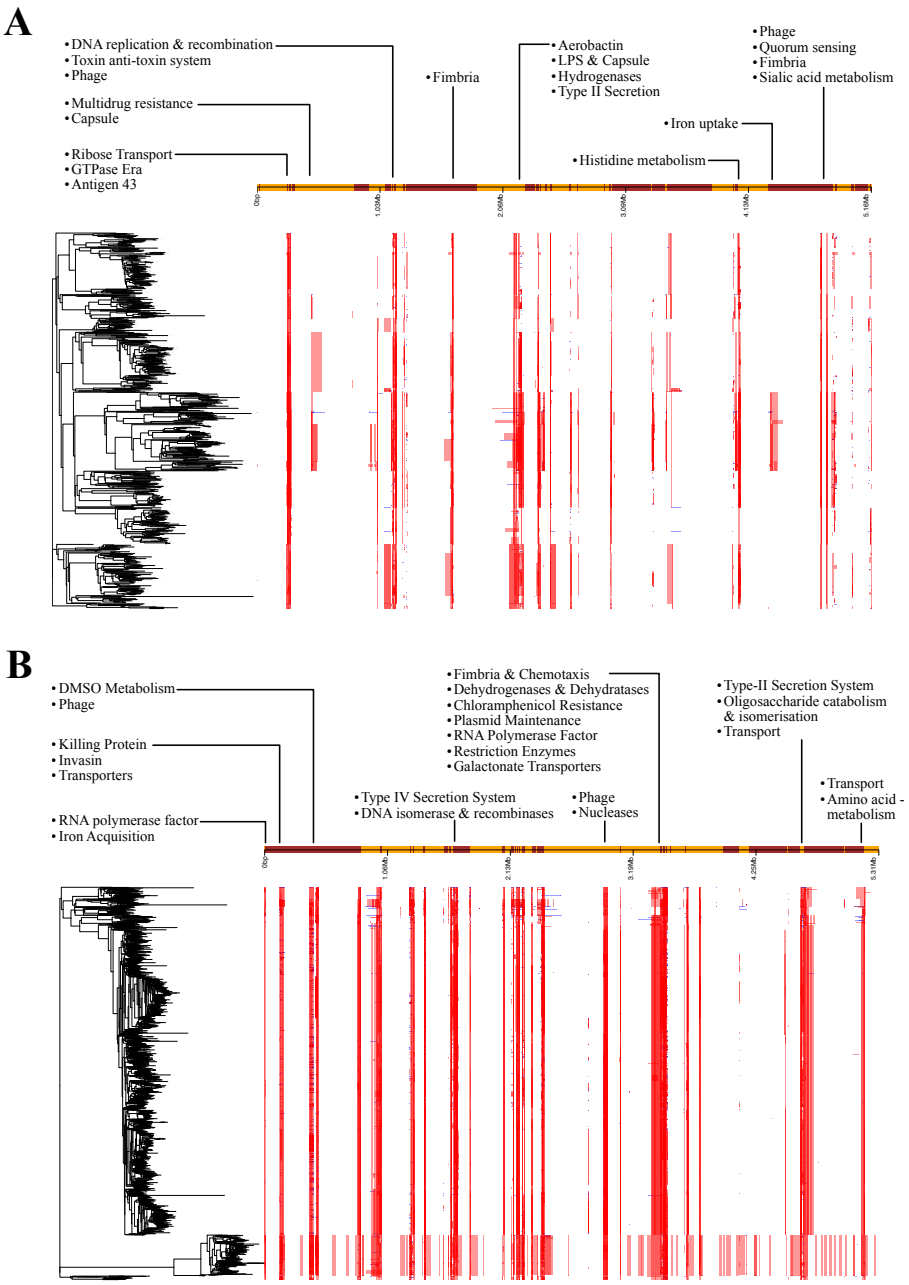

Fig. 3.20 Distribution of Recombination Blocks in the Genome of ST75 (A) and ST131 (B). Recombination predictions are indicated by red blocks. Reference genome contigs are indicated by alternating purle and yellow bocks. Phylogenetic tree produced by gubbins is in black. Selected gene functions are annotated.

3.3 Discussion

A dataset of approximately 19,000 *E. coli* genomes was assembled from Enterobase, which as of 2020, represents approximately 25% of all sequenced *E coli*. There is a strong bias towards human samples in the dataset as is expected given that there was a focus on populations of clinical relevance. Genomic identification of antibiotic resistance genes highlights the growing problem of antibiotic resistance. This is most clearly evident when examining the average number of resistance genes per genome; ST167, ST648, ST38 and ST131 all display on average in excess of 8 resistance genes per genome. This result is not driven by a large number of resistance genes within a small sub-population, as when examining the proportion of the lineage that is resistant it is clear that almost the entirety of these lineages possess multiple resistance genes. ST167, ST648, ST38 and ST131 all have in excess of 90% of their population possessing multiple resistance genes. This is in keeping with numerous other studies which frequently report these populations as major MDR pathogens [59, 67, 169, 116]. Moreover it confirms previous observations that AMR carriage is not equal across the population with AMR being concentrated in certain lineages [32, 3, 74].

Examining the pangenomes of these lineages reveals the huge amount of genetic diversity across *E. coli*. This is particularly evident in the ST10 and ST131 lineages. Unfortunately it is also apparent that there is a large amount of biological diversity that we have no functional information on as there are a large number of hypothetical protein annotations; as well as confirmed protein coding genes annotated with the 'Unknown Function' category. This problem has been encountered by others who found that the pangenome loci unique to ST131 were mostly hypothetical proteins [6], however other major categories they identified as unique to ST131 were flagellar and metabolic proteins. Previous pangenome analysis identified metabolic loci as being enriched in ST131 compared to other ExPEC, specifically anaerobic metabolic genes were exhibiting increased variation [96, 95]. Our functional pangenome analysis supports these observations, revealing that there is a significant

correlation between carriage of AMR genes and metabolism. Specifically there is increased variation in carbohydrate metabolism in populations with a high rate of antibiotic resistance carriage. Additionally these MDR lineages display increased levels of genes annotated with energy production and conversion in their core genome. Our GO term enrichment analysis identifies an increase in catabolic processes in MDR populations as well as variation in amino acid and sugar catabolism when compared to non-MDR populations. Previous analyses have focussed on individual lineages whereas here we present data on multiple MDR ExPEC lineages, revealing that metabolic variation is a shared adaptation. Together these point towards a strong role for metabolism in MDR lineage formation. Recent experimental evolution experiments have identified that *E. coli* can evolve resistance to antibiotic stress through mutations in core metabolic genes, particularly those involved in carbon and energy metabolism [84].

Metabolic screening assays have been employed to translate genomic observations to the lab. These assays have produced conflicting results with some observing that ST131 has a higher metabolic potential than other ExPEC and is capable of using a more diverse set of metabolites [55]; ST131 isolates that carried more AMR genes displayed a higher metabolic potential than ST131 isolates that lacked AMR genes. Subsequent studies using a larger screening library have contradicted this initial finding, instead observing that ST131 is not metabolically unique [4, 5]. Nor are ESBL positive *E. coli* more metabolically active than non-ESBL *E. coli* [5]. Instead, differences in metabolic capacity in ST131 isolates were strain specific and not shared across the lineage. Our pangenome analysis supports this observation as the correlation between carbohydrate metabolism and AMR carriage is restricted to the accessory pangenome, meaning that the allelic variants are not shared across the lineage and consequently individuals strains are unlikely to exhibit the same metabolic signature.

Whilst we have identified a significant association between variation in metabolic functions and carriage of antibiotic resistance we are unable to ascertain which occurred first. Genomic studies of ST131 determined that the acquisition of specific virulence factors preceded the acquisition of CTX-M-15 [11], implying that these virulence factors are important for the formation of MDR lineages. We attempted to apply this analysis to our dataset to determine if our observed association occurred before or after the acquisition of AMR genes. Specifically does the genetic change in metabolic categories permit carriage of antibiotic resistance or does it arise from adaptation to the presence of AMR genes. We attempted to do this using BEAST2 software [19], as was used to date the acquisition of specific virulence factors and AMR, however it became apparent that modelling the emergence of multiple genetic variants was not achievable. Instead we sought to identify the mechanism driving the observed genetic variation.

Multiple biological mechanisms were investigated in an attempt to identify the driving force behind the observed functional correlations. Previous work had identified that intergenic loci unique to ST131 and that these loci were undergoing allelic switching in clade C [96]. Moreover different CTX-M variants are associated with unique IGR profiles in the ST131 lineages [96], suggesting ST131 has adapted to AMR genes through acquisition of specific IGRs as has been observed in other bacteria [61]. Our analysis does not identify an association between total IGR diversity and carriage of AMR genes.

The levels of recombination in the selected populations were measured using Gubbins. The percentage of a genome in a recombination block was broadly similar across the majority of the population. The sole exception being ST10 which had a much higher proportion of the genome in a recombination block, confirming previous observations that K12 strains, which belong to ST10, display high levels of recombination in the core genome [94]. The levels of recombination observed do not correlate with resistance. Previous analysis identified that recombination in the core genome was reduced in ExPEC lineages and even more so in

ST131, leading the authors to propose that this is a hallmark of extra-intestinal pathogenicity [94]. We do not observe an association between pathogenicity and recombination, although our analysis here is limited to a single commensal lineage. Moreover the analysis presented here determined the rate of recombination across the entire genome whilst previous analysis focussed solely on the core genome [94]. Therefore our observations of a high level of recombination in ST131 could be reconciled with previous observations if ST131 undergoes low levels of recombination in the core genome but high levels of recombination in the accessory pangenome. In addition, analysis here was based upon lineages defined by MLST whilst previous recombination analysis was performed on genomes which had been clustered using Bayesian statistics which may result in more homogenous groupings of genomes compared to the approach used here [94].

In summary, the genomic analysis presented here indicates that MDR lineages of *E. coli* are displaying increased diversity in metabolic processes. Metabolism, particularly sugar and carbohydrate metabolism are known to be important for *E. coli* to colonise hosts. Specifically utilisation of carbohydrates are essential for commensal *E. coli* to colonise mice [26], importantly pathogenic *E. coli* utilise a different repertoire of carbohydrates compared to commensals [49]. This non-overlapping nutrient profile is thought to allow pathogenic *E. coli* to out-compete commensals, including other *E. coli*. Analysis presented here indicates that MDR lineages are displaying higher levels of variation than other pathogenic lineages, suggesting that MDR lineages may be further diversifying their metabolite repertoire. Does this altered repertoire of metabolites allow MDR lineages of *E. coli* to out-compete both commensal and non-MDR pathogenic *E. coli*? Previous investigations into the metabolic potential of ST131 have been conducted using high throughput screening assays. The growth conditions in these assays are very different from the host environment in which an invading MDR *E. coli* must survive. Importantly the conditions in which a bacteria grows can influence its metabolite usage, for example *S.* Typhimurium is capable of utilising the compound

tetrathionate however the importance of this is only apparent under competitive growth conditions. Moreover, the growth advantage conferred by tetrathionate usage is only apparent under anaerobic conditions [161]. Therefore, it is important to examine metabolic differences between commensal and MDR *E. coli* under competitive conditions, and to replicate the host environment.

Chapter 4

In Vitro Competition Suggests ST131 Out-competes Commensal *E. coli*

4.1 Introduction

ExPEC first colonise the gastro-intestinal tract of a host before escaping this compartment and disseminating to other sites where ExPEC cause disease. Transmission of ExPEC strains between hosts is via the faecal-oral route, with new hosts consuming food products contaminated with ExPEC. Upon entering a new host ExPEC are faced with the challenge of surviving and replicating in an environment that is already occupied by a multitude of commensal microbes. This entails competing with the resident microbes for a limited supply of nutrients. *E. coli* rely on carbohydrate metabolic genes to compete with other commensals to colonise the host, deletion of carobydrate utilisation genes from the commensal *E. coli* K12 MG1655 type strain impairs its ability to colonise mice [26]. Moreover pathogenic strains of *E. coli* use an altered repertoire of metabolites compared to commensal *E. coli* in order to out-compete them in colonising a host [49]. Our previous genomic analysis highlighted that carbohydrate metabolic genes and core energy production and conversion genes are exhibiting increased variation in MDR lineages of *E. coli*, suggesting that these lineages may have

further altered their metabolite utilisation profile. Previous metabolic screening assays have yielded contradictory results over the metabolic potential of ST131; with some indicating increased metabolic potential [56], whilst others observed that ST131 was not metabolically unique [4, 5]. However these assays do not replicate the environment encountered by ExPEC strains upon invading the host, where they must compete with other microbes. Mimicking this environment is incredibly important as it reveals differences that are otherwise obscured, for example competing strains of *S.* Typhimurium against one another only reveals a competitive difference under anaerobic conditions [161]. Examining the competitive ability for MDR and non-MDR isolates of *E. coli* is therefore of interest.

ExPEC are not the only pathogen to use the faecal oral route of transmission. *S.* Typhimurium is a highly successful enteric pathogen and faces the same challenge as ExPEC of establishing itself in the host gut whilst competing with native commensals. *S.* Typhimurium is capable of producing a bloom of growth in newly colonised hosts and so must be capable of efficiently outcompeting commensal microbes. Upon entering the host *S.* Typhimurium uses its virulence genes, such as type III secretion systems, to induce an inflammatory response from the host [29, 132]. Recruited host immune cells produce a variety of antimicrobial compounds including ROS [161]. The ROS then reacts with compounds present in the lumen of the gut thereby creating a new variety of metabolites [161]. *S.* Typhimurium is genetically equipped to utilise these novel compounds as terminal electron acceptors for anaerobic metabolism [161, 162]. The invading *Salmonella* has now created an ecological niche for itself by inducing the production of novel compounds that is it uniquely capable of exploiting. Other members of the Enterobacteriaceae family, including *E. coli*, have been observed to increase in abundance during the inflammatory conditions produced by *S.* Typhimurium as well as in autoinflammatory conditions [88, 141, 160, 41, 154]. Our previous results show that MDR lineages of *E. coli* display an increased variation in metabolic pathways. We therefore hypothesise that these genetic signatures are evidence that MDR *E. coli*

has evolved the same colonisation mechanism as *Salmonella*, specifically exploiting inflammatory metabolites and by-products.

To test these hypotheses, we will use an *in vitro* model of the gut epithelium to ascertain whether MDR *E. coli* induce a host inflammatory response. The Caco-2 and INT-407 cell lines will be used to model the host intestinal environment. Cell lines represent a highly simplified model of the host gut that does not incorporate the host immune system nor the native microbiome. The immune system is an important source of inflammatory mediators that can react with microbiome by-products to produce new metabolites utilised by invading pathogens. While the model adopted here only includes intestinal epithelial cells and omits immune cells, the epithelial cells represent the first line of defence for the host being responsible for interfacing with the microbiome and signalling the immune system. Epithelial cells are capable of modulating immune cells as well as producing pro-inflammatory cytokines such as IL-8, CXCL1 and TNFα [114]. Moreover a simplified model allows for rapid initial exploratory analysis which can be further expanded to include more complex components. To further test our hypothesis we will also compete commensal strains of *E. coli* against MDR *E. coli* to mimic the competition experienced by an invading *E. coli*. To measure the growth of both commensal and MDR strains simultaneously they will be differentially labelled with fluorescent protein encoding plasmids. Fluorescence can then be measured over time to determine if one strain dominates over the other.

4.2 Results

4.2.1 Competitive Growth on Eukaryotic Cell Monolayers

Multiple ST131 strains grow at a faster rate than K12 MG1655 on Caco-2 monolayer

Bacterial strains were transformed with plasmids encoding fluorescent proteins so that their growth could be measured in real time. For the purposes of assay development, the lab

strain K12 MG1655 was initially chosen as the commensal while multiple ST131 strains were used as the MDR competitor. The ST131 strains represented isolates from clades A, B and C. The ST131 strains were transformed with pMN402 which encodes GFP+ contollred by the heat shock protein 60 promoter (*hsp60*). Copy number of pMN402 is expected to be medium based on a ColE1 origin of replication. The K12 strain used had a chromosomeal insertion of mCherry at the *attB* lambda site. In order to mimic the intestinal environment bacteria were grown on monolayers of differentiated Caco-2 cells, a cell line derived from human colonic epithelial cells. These cells spontaneously differentiate once confluent to form polarised epithelial monolayers. Bacteria were inoculated onto the monolayers at a target multiplicity of infection (MOI) of 1. Fluorescent readings do not change for the first 300 minutes of culture before rising rapidly hitting the limit of detection in several instances (Figure 4.1a and b). The presence of the Caco-2 cells does not appear to affect the growth of the ST131 strains, however the K12 strain displays a reduction in the fluorescent signals when the monolayer is present. Optical density measurements show that bacteria achieve a higher density in the absence of Caco-2 cells. The strains assayed display variation in growth kinetics on Caco-2 cells with the ST131 clade A strains (F022 and F037) entering the log phase of growth between minute 200 and 300 whilst K12 only enters log phase after minute 500. When Caco-2 cells are absent all strains begin to enter log phase of growth at minute 200 but display different rates of growth with the ST131 clade A strains growing at the fastest rate whilst K12 exhibits the slowest rate. The fluorescent readings from multiple strains display a curious dynamic of increasing exponentially before peaking, decreasing then plateauing. This is not in keeping with the optical density measurements which display the standard sigmoidal growth kinetic. Moreover, the fluorescent measurements of strain F104 mimic those of K12 in the media only condition (Figure 4.1b) suggesting equivalent growth, yet the optical density measurements for these two strains are very different (Figure 4.1d). Additionally, the optical density measurements suggest that all strains reach an

equivalent level of growth at stationary phase yet this is not reflected in the fluorescent measurements. Ultimately this assay indicates that ST131 isolates grow better on monolayers of differentiated Caco-2 cells than K12.

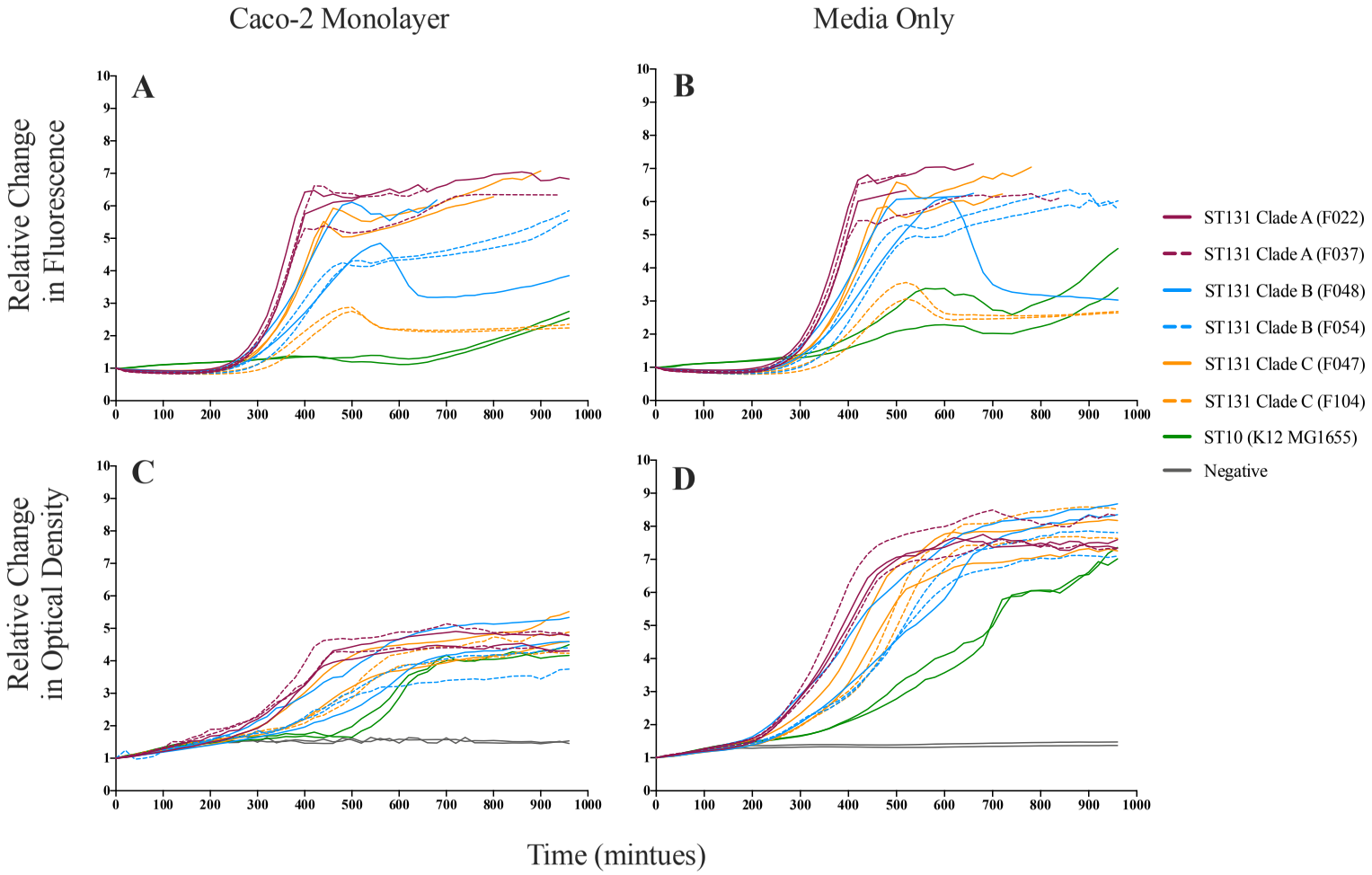

Fig. 4.1 Growth dynamics of ST131 strains and K12 on Caco-2 monolayers. Strains from ST131 clade A (maroon), B (blue), C (orange) as well as K12 MG1655 (green) were fluorescently labelled. Fluorescence (A & B) and optical density (C & D) were measured. Bacteria were grown on Caco-2 monolayers (A and C) as well as in DMEM only (B and D). Lines indicate invidual replicates, n=2.

Observing bacterial growth by time lapse microscopy suggests ST131 proliferates on differentiated Caco-2 monolayers whilst K12 growth is minimal

Bacteria were inoculated onto monolayers of differentiated Caco-2 cells at a target MOI of 1 and time lapse microscopy was conducted to directly visualise bacterial growth for 7 hours. Bacterial growth is readily visible as the background of Caco-2 cells is increasingly obscured by the growing bacteria (Figure 4.2). For the ST131 strain the eukaryotic monolayer is almost entirely obscured by 360 minutes. The K12 strain assayed displayed only minimal growth with the visualization of the Caco-2 monolayer uninterrupted for the duration of the imaging (Figure 4.3). Whilst the growth of ST131 was evident by phase contrast there was little change in the GFP channel for the first 360 minutes, it is only in the last 40 minutes of the experiment that GFP fluorescence becomes apparent (Figure 4.2). This is in keeping with the previous observation that optical density measurements start to increase exponentially after 200 minutes of culture but fluorescent readings remain stable till 300 minutes of culture. Despite the incongruence of and bacterial growth, it is evident from phase contrast that ST131 strains proliferated on Caco-2 monolayers whilst K12 growth was minimal.

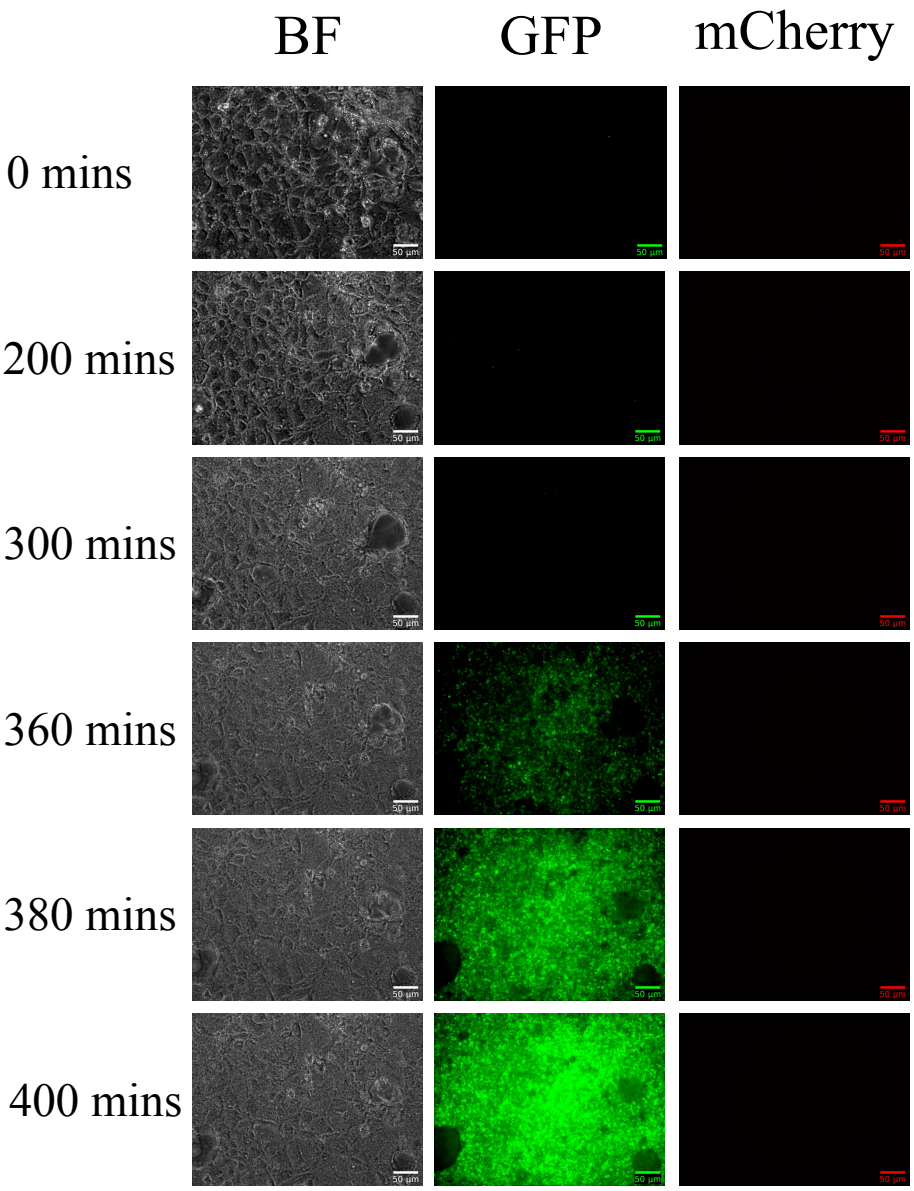

Fig. 4.2 Time lapse microscopy imaging of ST131 (F047 strain) on Caco-2 monolayers. The ST131 strain was transformed with a GFP reporter construct.

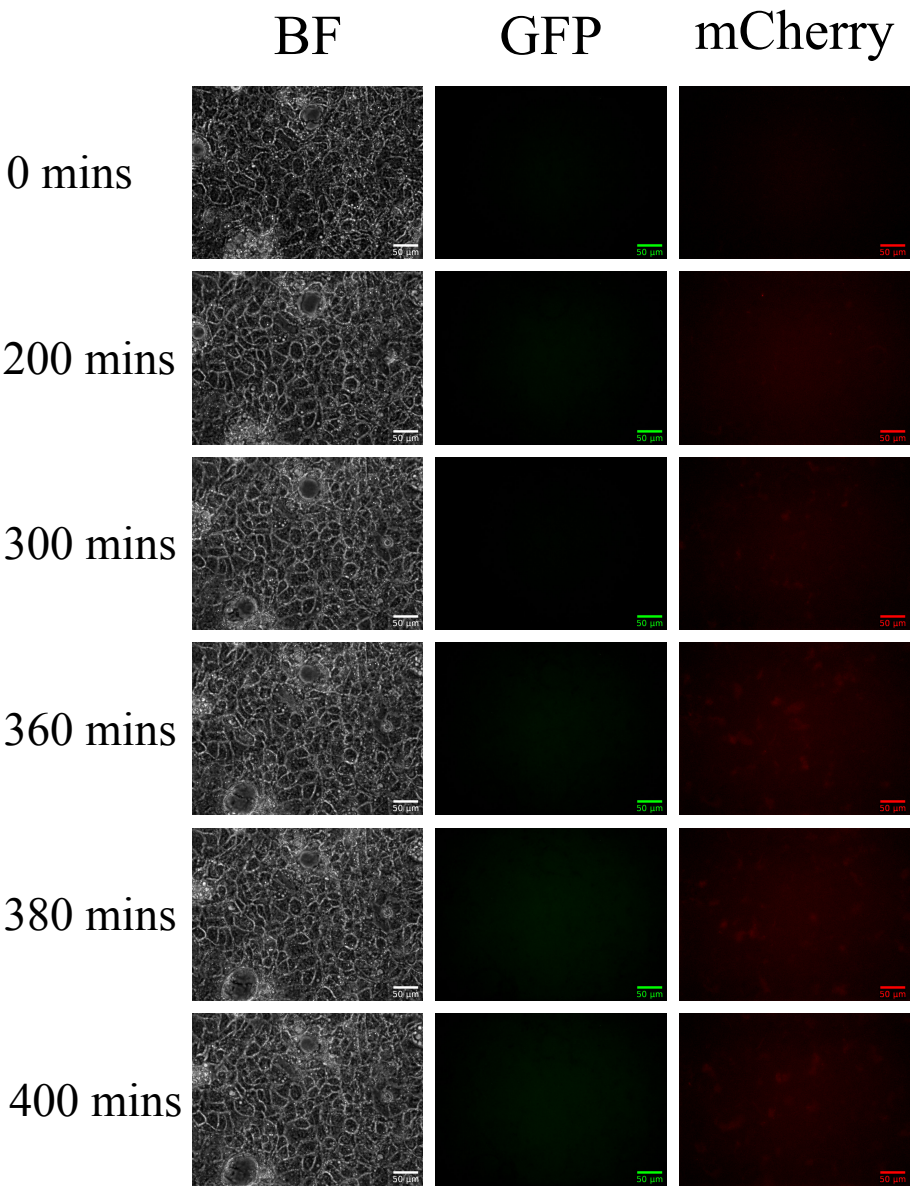

Fig. 4.3 Time lapse microscopy imaging of K12 on Caco-2 monolayers. The K12 strain was transformed with an mCherry reporter construct.

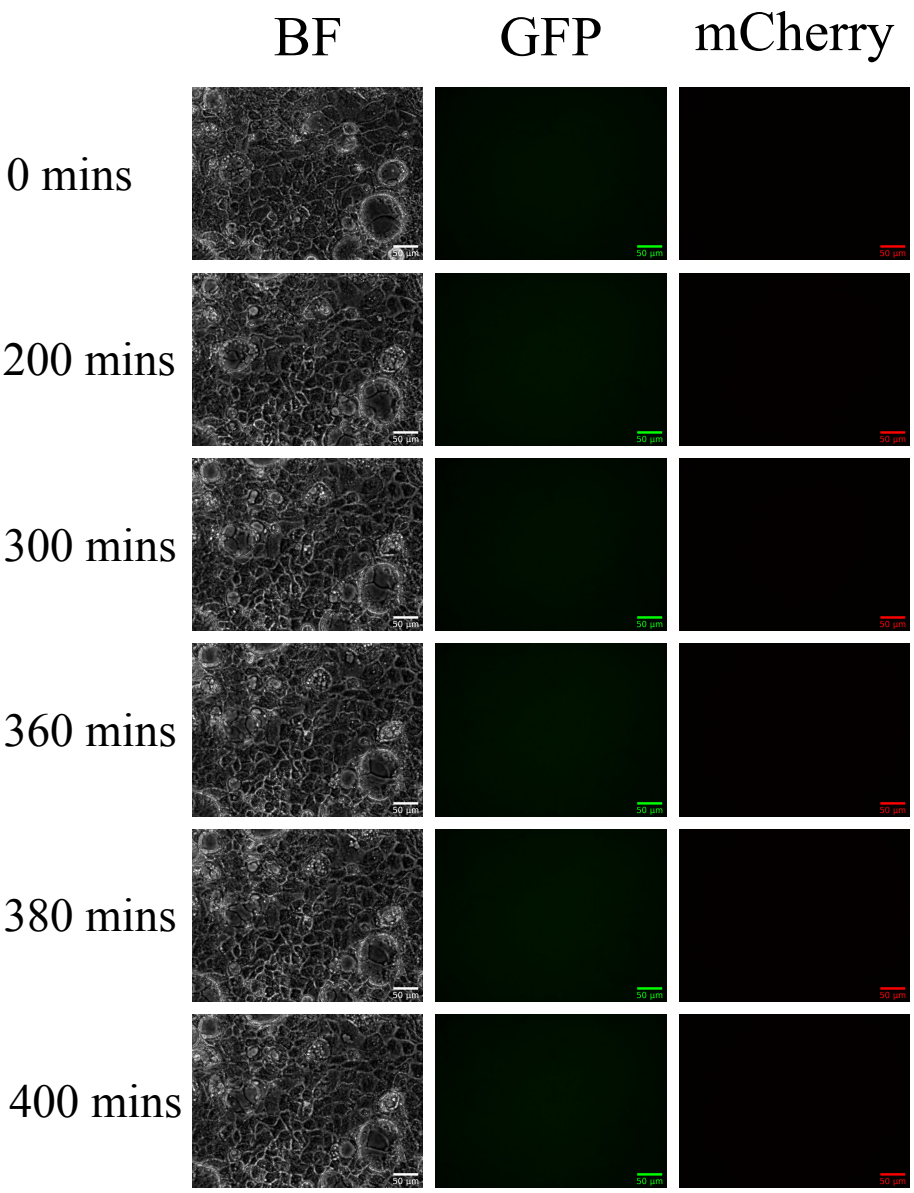

Fig. 4.4 Time lapse microscopy imaing of Caco-2 monolayers in the absence of bacteria.

Presence of ST131 strain suppresses fluorescence signal from K12 strain on INT407 monolayers

In order to further explore if our hypothesis holds promise, new fluorescence plasmids were transformed into the bacteria as observations from previous experiments suggested that fluorescence was not congruent with bacterial growth. The plasmids pCherry8 and pFPV25.1 were acquired which encode mCherry and GFP-mut3a respectively. Expression of the fluorescent protein is driven by *rpsA* in pCherry8 and *rpsM* in pFPV25.1, both promoters are constitutive being derived from ribosomal protein genes. While the copy number of these constructs has not been determined the plasmid backbone of pCherry8 (pSMT3-M) is a high copy number plasmid while the backbone of pFPV25.1 (pFPV) is a low copy number plasmid. For the purposes of assay optimisation, the Caco-2 cells used previously were replaced by the faster growing INT-407 cell line, which is also derived from human intestinal epithelial cells. When K12 was grown in isolation on the INT-407 monolayer there is a clear GFP signal detected that better mimics the measurements for optical density (Figure 4.5). Both fluorescent signal and optical density start to increase after 100 minutes of culture. The ST131 clade C strain (F104) grown in isolation on the monolayer appears to grow at a slightly faster rate than K12 when measured by optical density (Figure 4.5c). Unfortunately, there is little change in mCherry fluorescence over the course of the experiment with a minor increase in fluorescence only observed once the bacteria have reached stationary phase (Figure 4.5b). Suggesting that again the expression of the fluorescent protein is not constitutive and cannot be used to infer bacterial growth. When the two strains are mixed together there is a clear suppression of the GFP signal from the K12 strain suggesting its growth is impeded by the presence of ST131 (Figure 4.5a).

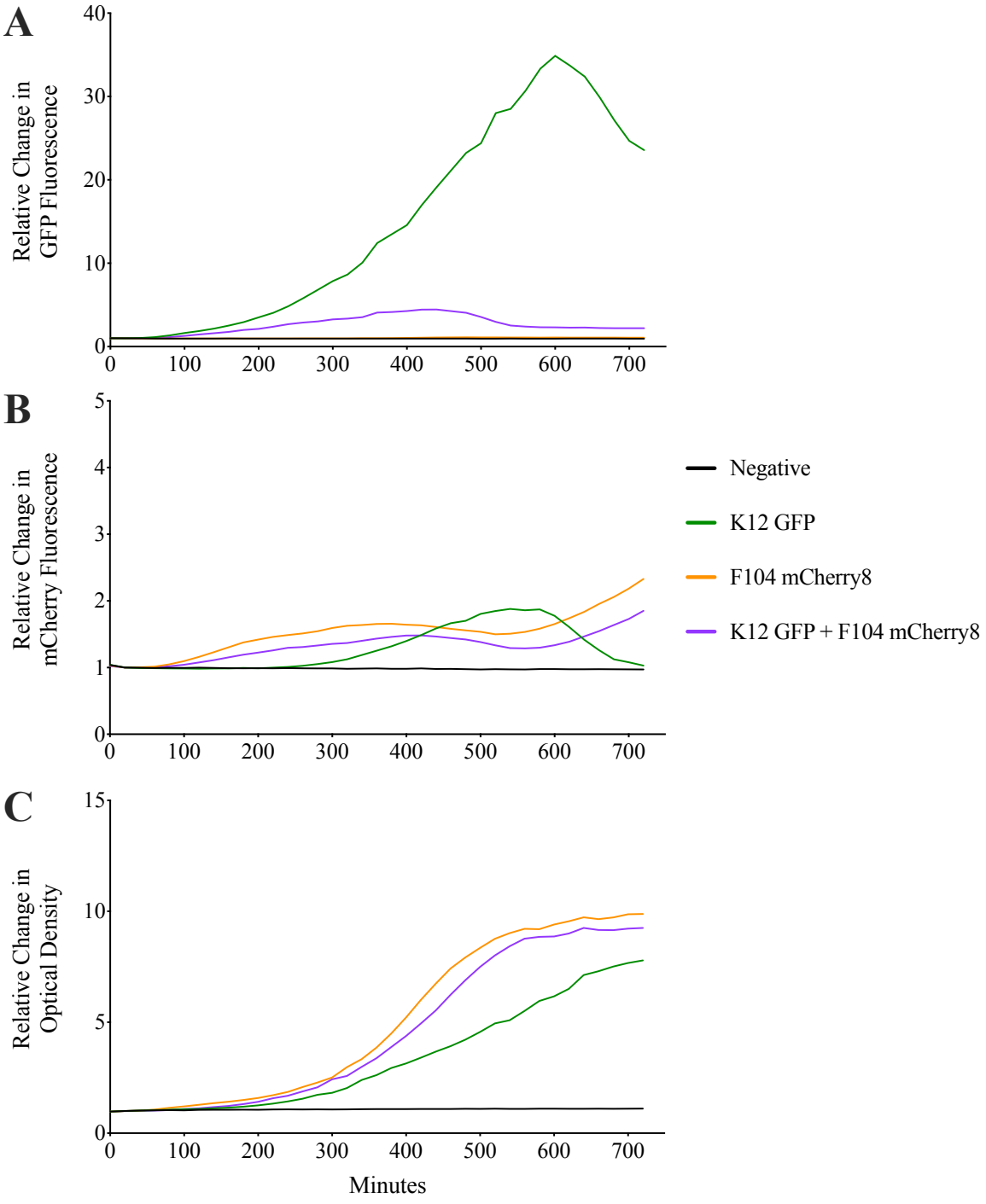

Fig. 4.5 Growth dynamics of strains on INT407 cells. Bacteria were grown independently on INT407 cells (ST131 mCherry: orange line, K12 GFP: green line) as well as in a 1 to 1 mixture (purple line). A negative control in which no bacteria were inoculated (black line) was included. GFP fluorescence (A), mCherry fluorescence (B) and optical density (C) were measured. Lines indicate invidual replicates, n=1.

Suppression of fluorescence from K12 strain by ST131 strain does not require the presence of INT407 monolayer

Fluorescence was also measured using flow cytometry allowing individual cells to be enumerated. Bacteria were grown on INT407 monolayers at an MOI of 1 for 6 hours before resuspension and analysis on a flow cytometer. Again, GFP fluorescence was detected in K12 bacteria when the strain was grown individually (Figure 4.6d). When ST131 (F104) and K12 were cultured together in a 1 to 1 ratio there was a suppression of the fluorescence from K12, suggesting that the presence of the ST131 strain impaired the growth of K12. This effect occurred both in the presence and absence of INT407, suggesting that the growth media used could be causing the difference. To confirm this the bacteria were grown in combination with two different media (LB: nutrient rich and DMEM: minimal media) and growth was measured in a plate reader. Fluorescence from K12 was detected in both LB and DMEM media and accumulated at similar rates (Figure 4.7a). Growing ST131 (F104) in LB produced an improvement on the fluorescent signal but it was still delayed to late log phase of growth (Figure 4.7b). By optical density both strains grew slightly better in LB than in DMEM (Figure 4.7c). When the two strains were grown together in a 1 to 1 ratio in LB there was minimal suppression of the GFP signal from K12 whilst when grown in DMEM there is a clear reduction in the signal. From fluorescent measurements we can infer that the growth of K12 is supressed in presence of ST131 in nutrient limited media.

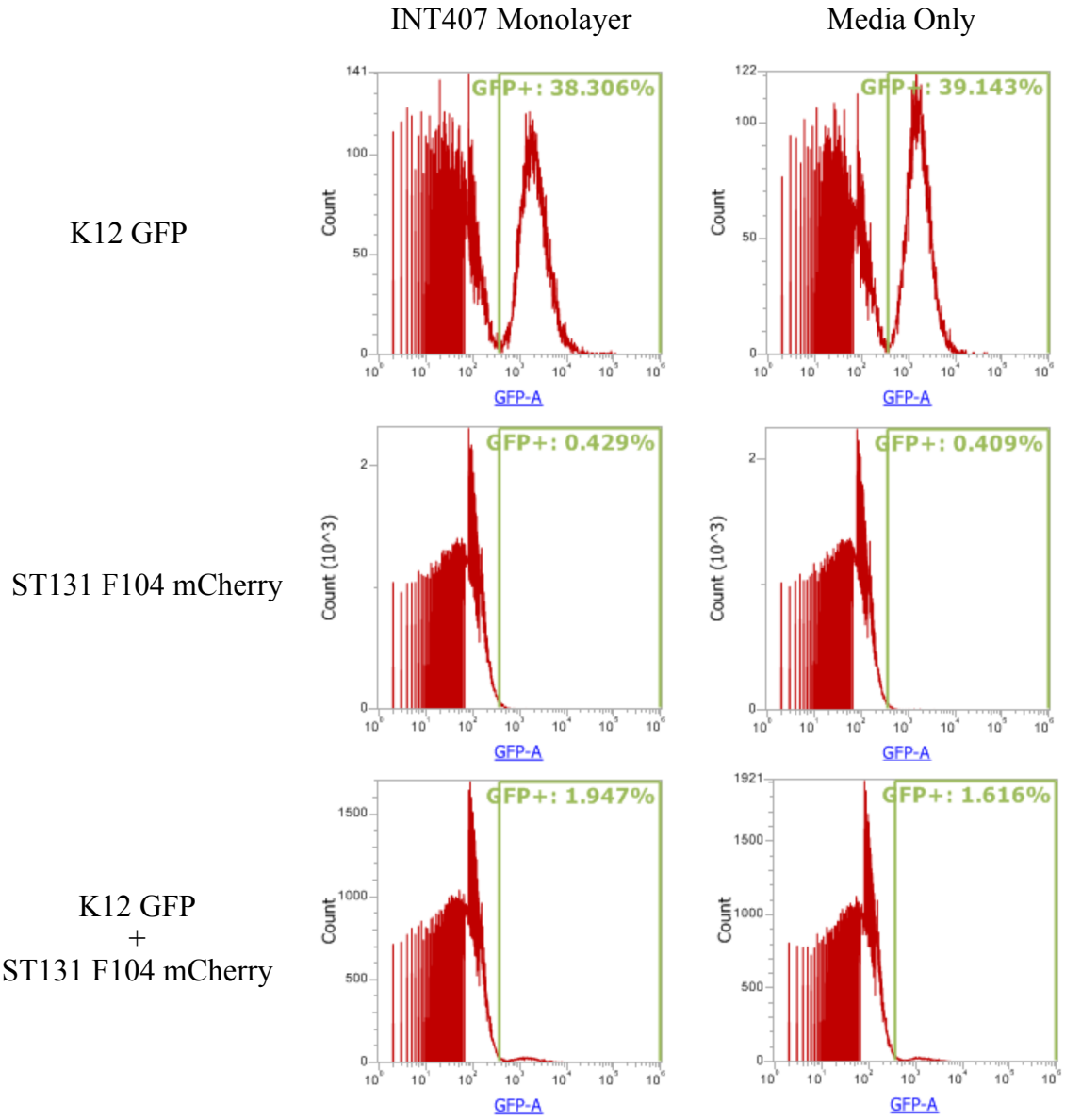

Fig. 4.6 Bacterial fluorescence measured by flow cytometry after culture on INT407 cells. Bacteria were grown on INT407 monolayers or in cell culture media only. Bacterial strains were added to monolayers both separately and in a 1 to 1 mixture. GFP positive cells are readily identifiable when K12 GFP is analysed. A minimal number of GFP positive cells are detected when K12 is absent from the culture. When both strains are mixed the number of GFP positive cells remains minimal, both in the presence and absence of INT407 cells.

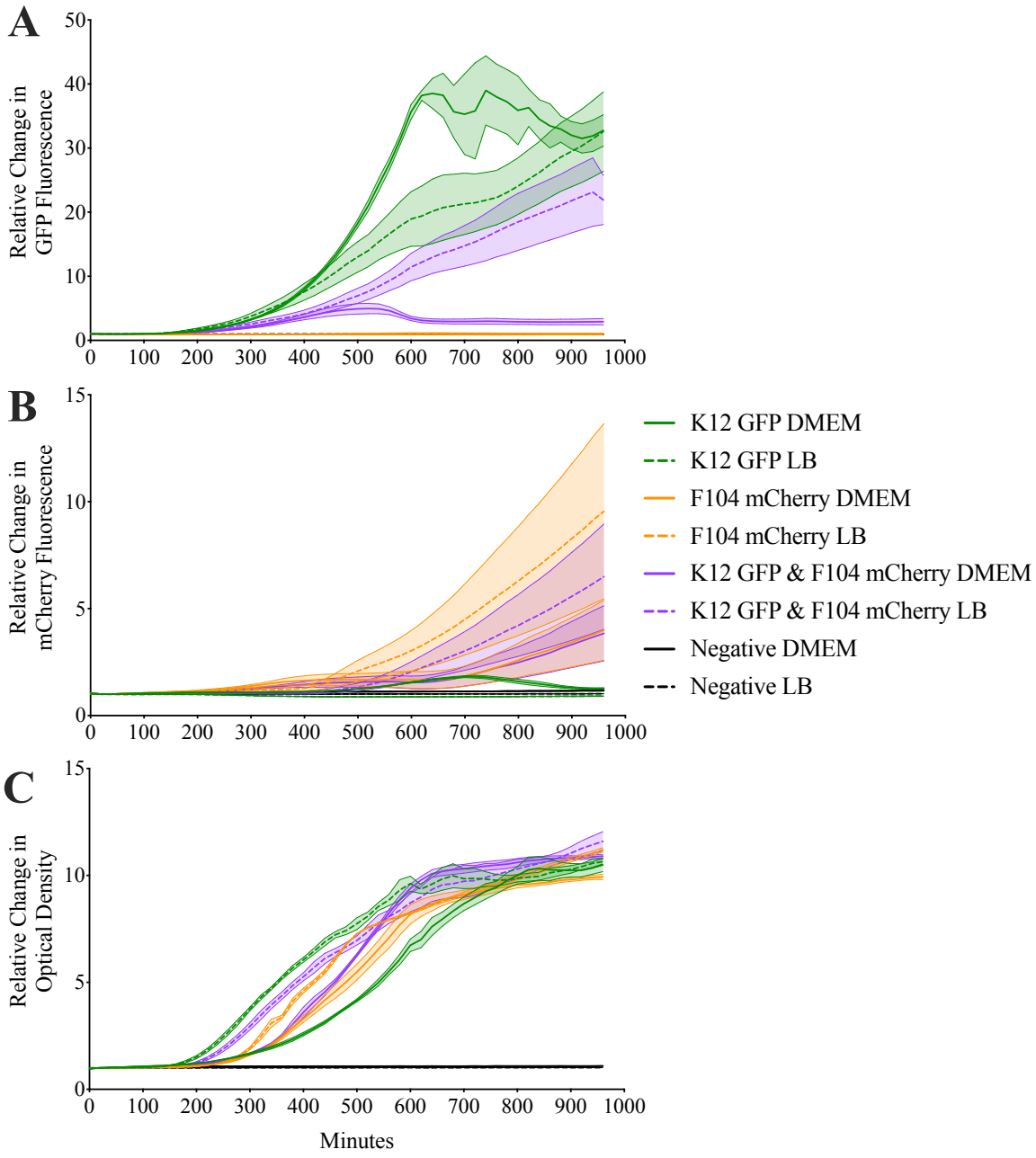

Fig. 4.7 Growth dynamics of strains in LB and DMEM media. Bacteria were grown in LB (dashed) and DMEM (solid). The K12 GFP strain was grown in isolation (green) as was the ST131 mCherry strain (orange). The strains were also mixed in a 1 to 1 ratio (purple). GFP fluorescence (A), mCherry fluorescence (B) and optical density (C) were measured. Bold lines indicate average while shaded areas indicate SEM, n=3.

4.2.2 Inflammatory Response from Cell Lines

Both Caco-2 monolayers and INT407 monolayers do not produce a robust inflammatory cytokine response in the presence of ST131

We hypothesise that MDR *E. coli* generate an inflammatory response from the host in order to aid its ability to colonise. We expect the inflammatory response to be greater than for non-MDR commensal *E. coli*. Previous work has shown that Caco-2 monolayers produce TNFα, IL-6 and IL-8 in response to bacteria whilst INT407 monolayers produce IL-1α, IL-1β, TNFα, IL-6, IL-98 and MCP-1 [114, 8]. To confirm whether the eukaryotic cell monolayers were responding to the presence of our bacterial strains a multi-analyte Enzyme-linked Immunosorbent Assay (ELISA) was conducted. Bacteria were grown on the cell monolayers for 6 hours before the culture supernatant was collected for the assay. The monolayers lack a robust cytokine response in the presence of bacteria with the sole exception of K12 which elicits a strong IL-8 response from Caco-2 cells (Figure 4.8). Overall the ST131 strains used elicit a moderate IL-8 response from Caco-2 cells. The K12 strain also elicits a minor MCP1 and MIP1b response from the Caco-2 cells. The other cytokines assayed appear unaffected by the presence of bacteria. Additionally, the Lipopolysaccharide (LPS) used as a positive control does not produce a cytokine response, suggesting that the Caco-2 cells are sensing the bacteria via other Pathogen Associated Molecular Patterns (PAMPs). The INT407 cells appear to be constitutively expressing both IL-8 and RANTES and their expression is not affected by the presence of bacteria. Moreover, bacteria do not alter the expression of any cytokines from INT407 cells.

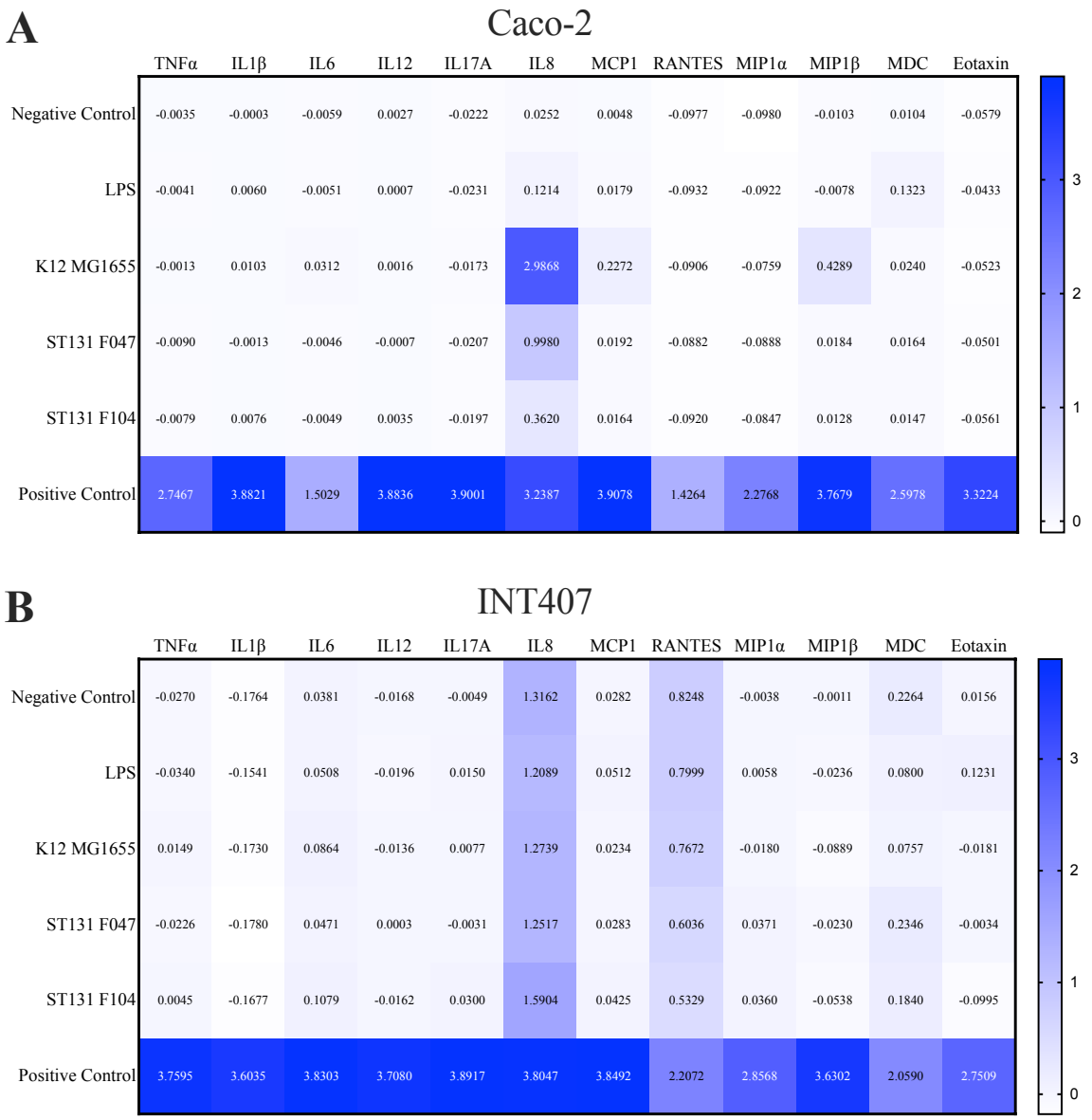

Fig. 4.8 Multi-analyte ELISA heatmap. Multiple strains (rows) were cultured on Caco-2 (A) or INT407 (B) for 6 hours before supernatant was collected. Multiple cytokines (columns) were analysed. Hue of colour indicates strength of signal, a darker blue indicates a higher concentration of cytokine while white indicates absence. Purified cytokine was used as an internal positive control to confirm assay reagents. LPS was used as an external positive control to confirm cells were capable of producing cytokines. The negative control was taken from untreated cells to determine basal levels of cytokine production.

4.2.3 Bacterial Growth in Nutrient Rich and Poor Media

ST131 strain is able to out-compete commensal strain in both nutrient rich and poor media

To attempt to confirm our observations of ST131 negatively affecting K12 growth a new method for measuring bacterial growth was used. As before bacteria were grown either in isolation or as a 1 to 1 mixture for 6 hours before diluting and plating onto agar plates to enumerate Colony Forming Unitss (CFUs). Total bacterial growth was measured as well as growth on selective antibiotics, allowing the growth of the MDR strains to be distinguished. Again, competition between the strains was measured in both LB and DMEM. Additionally, an ST73 (F084) was included in the assays. ST73 shares the same ExPEC pathogenic phenotype as ST131, with both isolates coming from patients with blood stream infections, however ST73 is rarely identified as MDR. Our results from Chapter 3 demonstrate that metabolic adaptation is correlated with carriage of AMR, therefore we predict that whilst ST73 shares a pathogenic phenotype with ST131 it does not share its metabolic capabilities and consequently will behave differently.

In LB all strains assayed grew to the same average CFU: ST131 (F016) growing to 4.63×10^9 CFU/ml; ST73 (F084 growing to 4.49×10^9 CFU/ml; ST10 (811E7) growing to 4.02×10^9 CFU/ml (Figure 4.9). In DMEM the ST131 strain grew to the highest CFU (2.22×10^8 CFU/ml) followed by ST73 which grew to near identical levels (1.90×10^8 CFU/ml) whereas the growth of the commensal ST10 was 10-fold lower than the other strains (9.55×10^6 CFU/ml). These observations are consistent with the previous growth kinetic measurements showing that the commensal strain grew at a slower rate than ST131. Strains grown competitively can be distinguished based on their resistance profiles, with the ST10 being pan-susceptible, the ST73 being resistant to sulfamethoxazole and the ST131 being resistant to both sulfamethoxazole and ciprofloxacin. Total growth of bacterial competition assays was measured from growth on non-selective agar whilst ExPEC strains were distin-

guihsed from the commensal by growth on selective agar. When competing ST10 against ST131 in LB 72.4% of the total growth was resistant to ciprofloxacin (4.10×10^9 CFU/ml average total growth, 2.97×10^9 CFU/ml average resistant growth). When competing ST10 and ST131 in DMEM even more of the growth is attributable to ST131 with 89.9% of total growth being resistant to ciprofloxacin (8.39×10^7 CFU/ml total, 7.54×10^7 CFU/ml resistant). Competing ST10 against ST73 in LB resulted in 68.3% of the total growth being resistant to sulfamethoxazole (3.94×10^9 CFU/ml total, 2.69×10^9 CFU/ml resistant). When these strains are grown in DMEM ST73 is responsible for 61.8% of growth (1.86×10^8 CFU/ml total, 1.15×10^8 CFU/ml resistant). Additionally, ST73 and ST131 were competed against one another in LB where ST131 is was responsible for 40.0% of growth (4.59×10^9 CFU/ml total, 1.79×10^9 CFU/ml resistant). In DMEM ST131 was responsible for 35.5% of growth (3.01×10^8 CFU/ml total, 1.07×10^8 CFU/ml resistant). These results are consistent with our previous observations that ST131 out-competes the commensal.

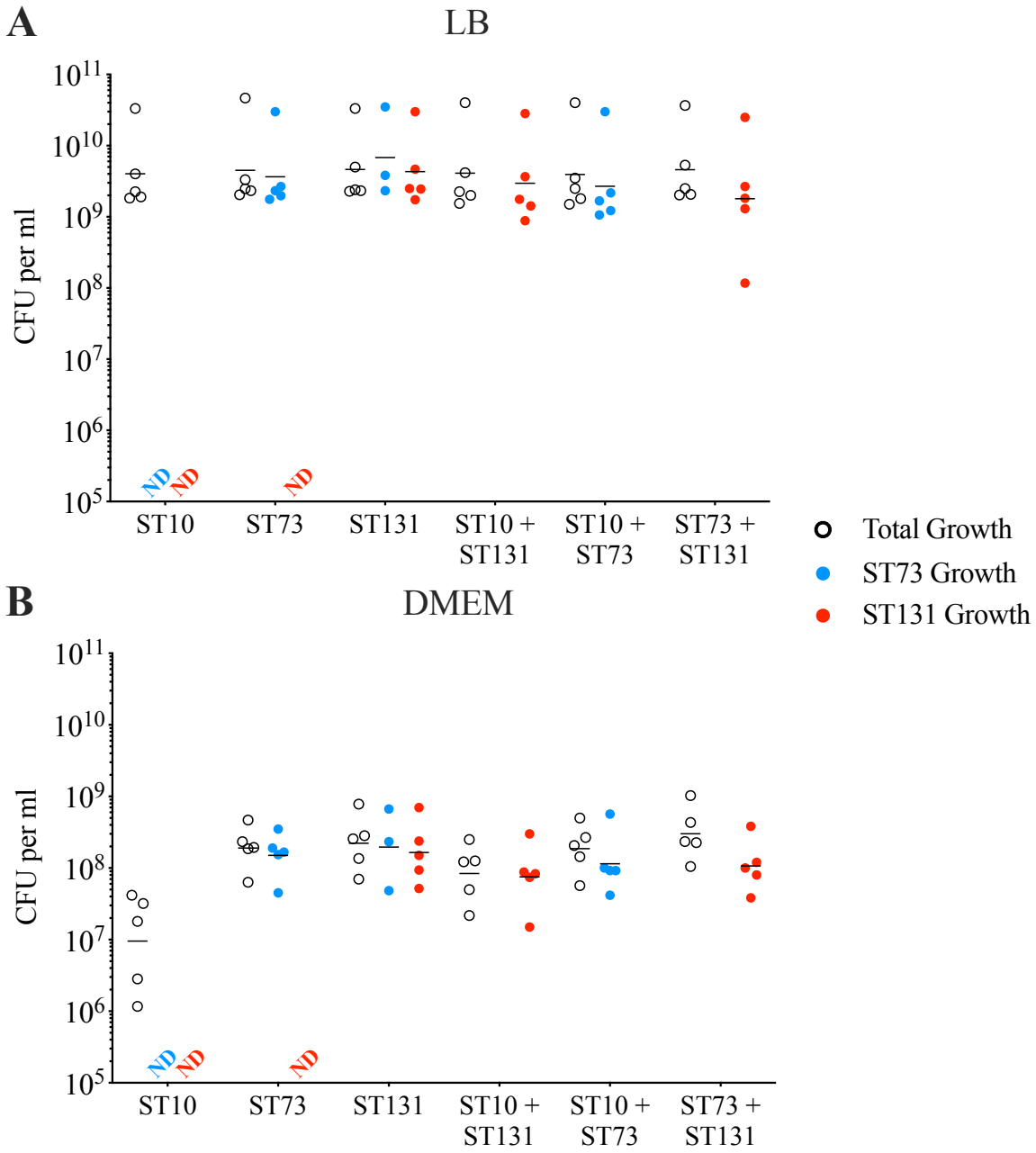

Fig. 4.9 Growth of multiple strains in LB and DMEM as determined by CFU assay. A commensal ST10 (811E7), ST73 (F084) and ST131 (F016) were grown in isolation or in competition in LB (A) or DMEM (B) for 6 hours. Bacteria were competed in a 1 to 1 ratio. Total growth was measured on non-selective agar (open circles), ST73 growth was determined by sulfamethoxazole resistance (blue) and ST131 growth was determined by ciprofloxacin resistance (red). Growth of ST10 was not detected on sulfamethoxazole (blue ND) nor ciprofloxacin (red ND). Growth of ST73 was not dected on ciprofloxacin (red ND). Lines indicate geometric mean, n=5.

ST131 strain maintains competitive advantage over commensal under hypoxic conditions

Previous work identified that anaerobic metabolic genes had an increased abundance in ST131 compared to other ExPEC lineages [95]. We therefore hypothesised that the competitive advantage exhibited by ST131 would be increased in oxygen limited environments. The previous CFU assay was repeated but with an oxygen concentration of 5%. The total growth of all strains in LB was decreased by a log fold when oxygen concentrations were reduced. The commensal ST10 grew to 9.43×10^8 CFU/ml, the ST131 to 8.09×10^8 CFU/ml and the ST73 to 8.03×10^8 CFU/ml (Figure 4.10, Table 4.1). Conversely, reduced oxygen concentrations increased the total growth measured in DMEM and were comparable to the total growth measured in LB. ST73 grew to 6.33×10^8 CFU/ml, ST131 to 6.03×10^8 CFU/ml and ST10 to 1.7×10^8 CFU/ml. Reducing the oxygen concentration improved the growth of the commensal strain by a log fold. Again, the strains were competed in pairs in both media types at the lower oxygen concentration. Competing ST10 against ST73 showed that ST73 produced the majority of the growth with 66.17% of the total growth being resistant to sulfamethoxazole (7.98×10^8 CFU/ml total , 5.28×10^8 CFU/ml resistant). When grown in nutrient limited DMEM ST73 was responsible for 51.73% of total growth (5.49×10^8 CFU/ml total, 2.84×10^8 CFU/ml resistant). The ST10 strain was also competed against ST131 in LB in which ST131 accounted for 61.97% of the growth (8.02×10^8 CFU/ml total, 4.97×10^8 CFU/ml resistant). While in DMEM media the ST131 out-competed the commensal accounting for 72.96% of growth (3.55×10^8 CFU/ml total, 2.59×10^8 CFU/ml resistant). Finally, ST73 was competed against ST131 in which neither strain dominated over the other in either LB (52.87% ST131 growth, 5.92×10^8 CFU/ml total, 3.13×10^8 CFU/ml resistant) or DMEM (52.16% ST131 growth, 6.94×10^8 CFU/ml total, 3.62×10^8 CFU/ml resistant). This demonstrates that the ST131 strain is able to out-compete the com-

mensal strain under all conditions but performs best in the nutrient limited DMEM media under aerobic conditions.

Table 4.1 Geometric mean of CFU values for strains grown under multiple conditions. Averages were calculated from five replicates for experiments at 20% oxygen and four replicates for experiments at 5% oxygen.

	LB Broth		DMEM Media	
Strain	20% Oxygen	5% Oxygen	20% Oxygen	5% Oxygen
ST10 (811E7)	4.02 x10^9	9.43 x10^8	9.55 x10^6	1.70 x10^8
ST73 (F084)	4.49 x10^9	8.03 x10^8	1.90 x10^8	6.33 x10^8
ST131 (F016)	4.63 x10^9	8.09 x10^8	2.22 x10^8	6.03 x10^8

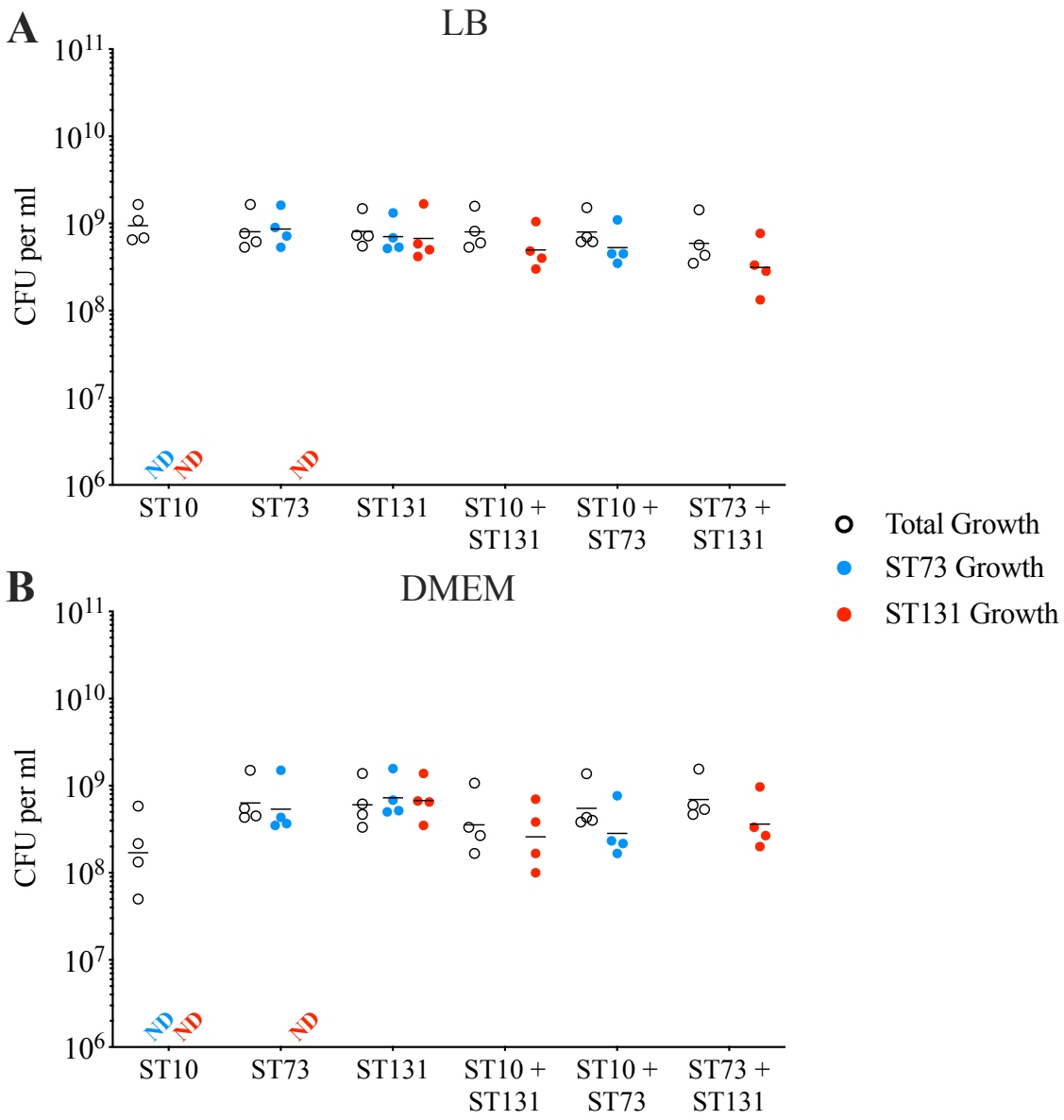

Fig. 4.10 Growth of multiple strains in LB and DMEM at 5% oxygen as determined by CFU assay. A commensal ST10 (811E7), ST73 (F084) and ST131 (F016) were grown in isolation or in competition in LB (A) or DMEM (B) for 6 hours. Bacteria were competed in a 1 to 1 ratio. Total growth was measured on non-selective agar (open circles), ST73 growth was determined by sulfamethoxazole resistance (blue) and ST131 growth was determined by ciprofloxacin resistance (red). Growth of ST10 was not detected on sulfamethoxazole (blue ND) nor ciprofloxacin (red ND). Growth of ST73 was not dected on ciprofloxacin (red ND). Lines indicate geometric mean, n=4.

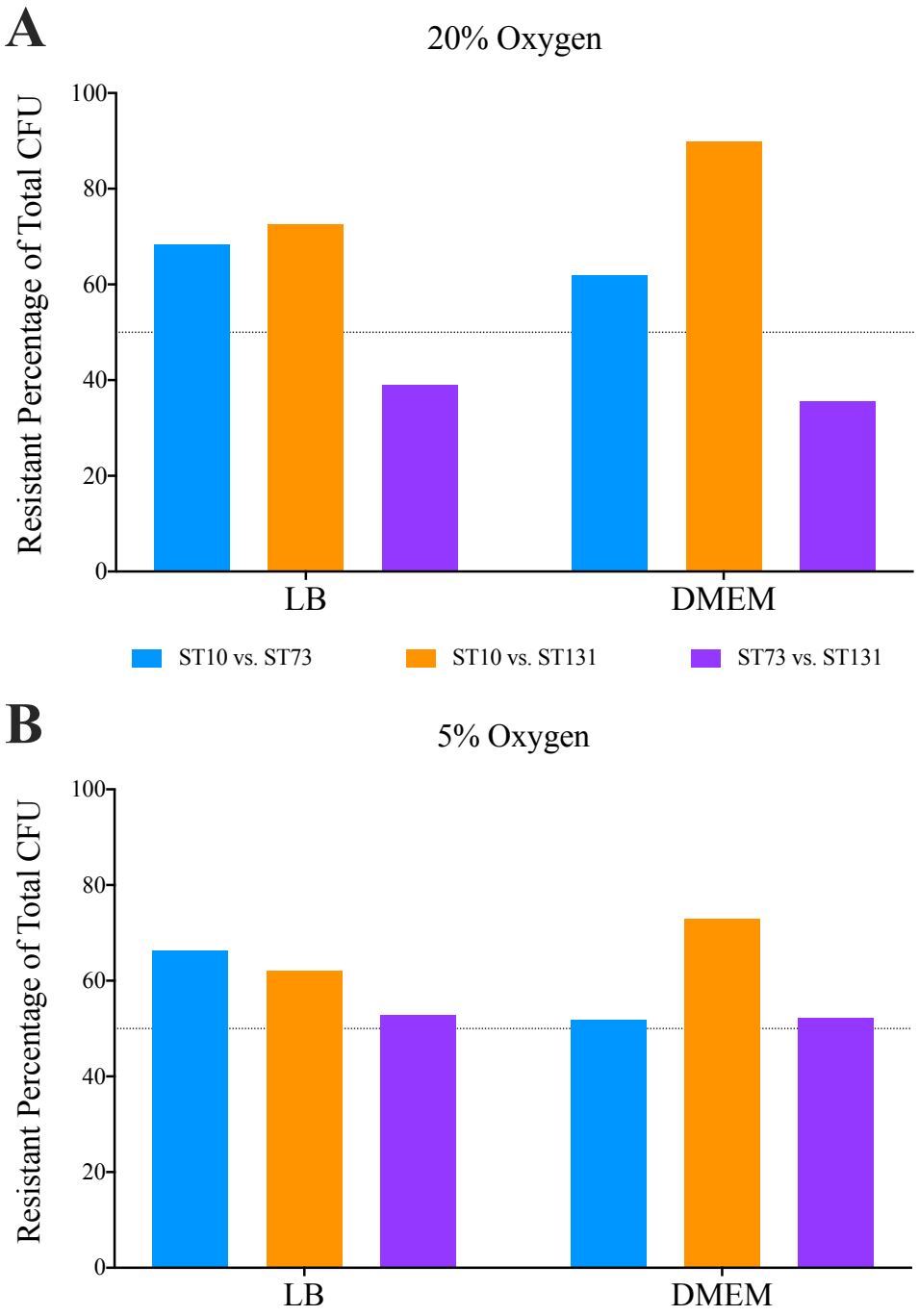

Fig. 4.11 Percentage of growth attributable to the resistant strain, summary data from Figure 4.9 (A) and Figure 4.10 (B). Bars indicate the percentage of average growth that is resulting from the resistant strain, either ST73 and ST131.

Observed effect on K12 growth is not caused by a secreted substance or phage

The previously observed ability for ST73 and ST131 to out-compete the commensal strain could be a result of these strains actively killing their competition. This could be achieved by either the release of a lytic phage, a diffusible toxin or type VI secretion system. To test whether bacterial killing was occurring bacterial cultures were grown to mid-log phase before harvesting the culture supernatant. The supernatant was filter sterilised before mixing with an equal volume of fresh LB media. Strains were then inoculated into the diluted supernatant and growth was measured by optical density readings. The growth of the commensal ST10 was not affected by the supernatant of ST73 nor ST131 as it displayed the same growth dynamics as when it was grown in its own supernatant (Figure 4.12). The same is true for both ST73 and ST131. The filtered supernatant was sterile as no growth was detected when no bacteria were inoculated. This indicates that the assayed strains do not kill each other using a diffusible toxin or lytic phage.

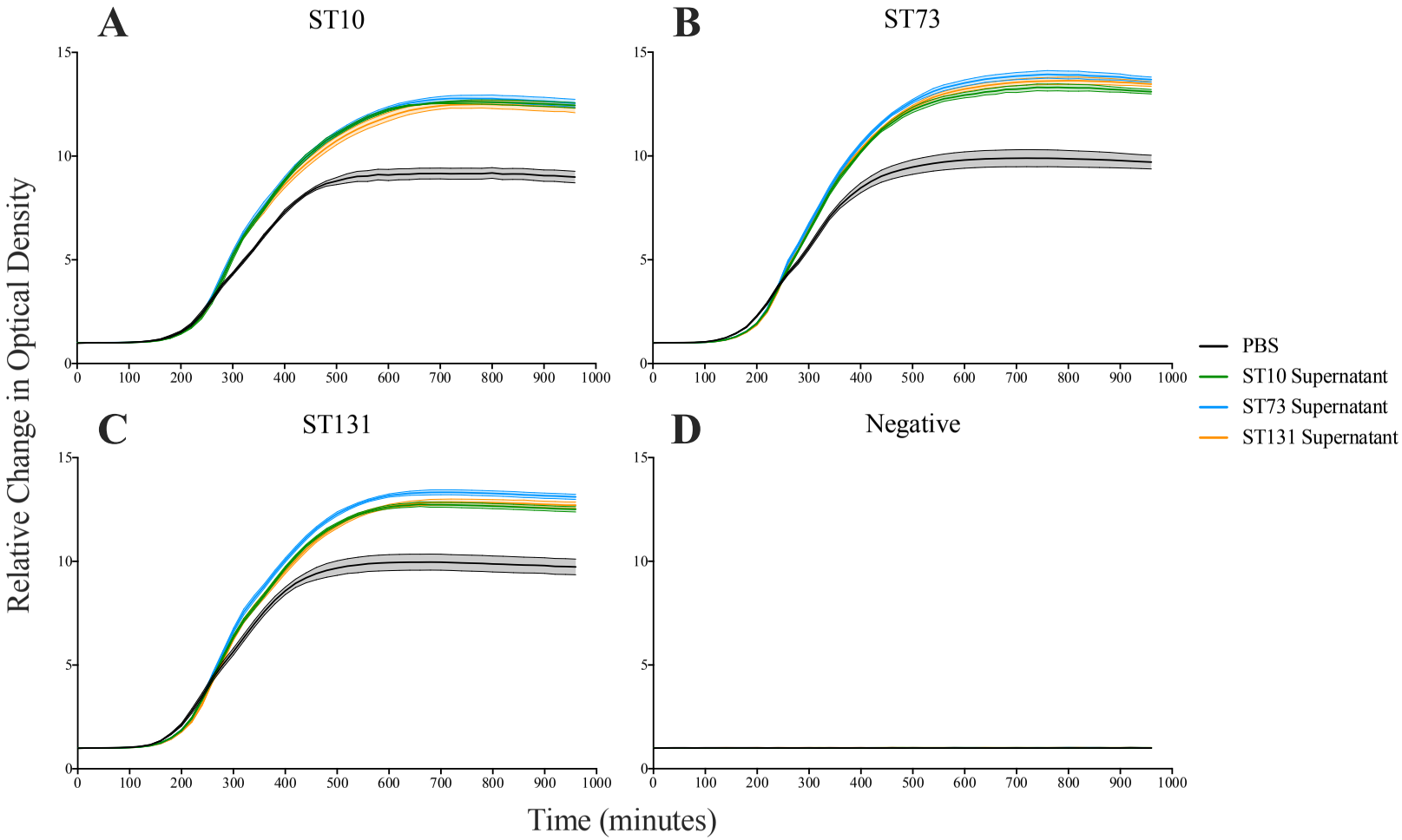

Fig. 4.12 Growth of strains in diluted culture supernatants. The ST10 strain (811E7, panel A), the ST73 strain (F084, panel B) and the ST131 strain (F016, panel C) were grown in 50% LB with 50% culture supernatant from ST10 (green line), ST73 (blue line) or ST131 (orange line). Strains were also grown in 50% LB supplemented with 50% PBS (black line). No growth was observed when no bacteria were inoculated (D). Bold lines indicate mean while shaded areas are SEM, n=5.

4.3 Discussion

Previous genomic analysis has identified that ST131 displays increased variation in anaerobic metabolic genes. The work presented in Chapter 3 expands on this work to identify that multiple MDR ExPEC lineages display increased variation in carbohydrate metabolism and increased abundance in core metabolic pathways compared to non-MDR lineages. Pathogens invading a new host must compete with the resident commensals for scarce nutrients, the invading strains must be able to out-compete the residents in order to colonise the host. For example a pathogenic strain of *E. coli* (O157:H7) competes with commensal *E. coli* strains for a select set of sugars, if all the sugars required by the pathogenic *E. coli* are consumed by commensals the pathogen cannot establish itself [90]. Here we propose that our observed metabolic signature in MDR lineages is an adaptation employed by MDR lineages to excel at metabolic competition with commensals. Previous work has focussed on determining the metabolic potential of ST131 resulting in conflicting conclusions. Initially ST131 was reported to exhibit increased metabolic potential [55], whilst subsequent studies demonstrated that ST131 did not exhibit unique metabolic characteristics [4, 5]. Here we present data on bacterial competition rather than determining the metabolic potential of strains in isolation, moreover we employ intestinal epithelial cells to mimic the host environment. Recreating the conditions that ExPEC face in their natural environment is important as has been demonstrated for *S.* Typhimurium. Specifically *S.* Typhimurium is capable of using tetrathionate however the growth advantage that this confers is only apparent under anaerobic conditions [161]. Our data indicate that under more natural conditions ST131 is capable of out-competing commensal strains of *E. coli*. This is in keeping with observations that ST131 is able to easily spread between individuals [98, 69], implying that it is adept at out-competing commensal microbes. The use of human intestinal epithelial cells to mimic the presence of the host did not affect bacterial competition, but did limit the total amount of bacterial growth achieved. Instead it became apparent that the culture media in

use could influence bacterial competition, with media in which nutrients are limiting conferring the greatest advantage to ST131 over a commensal. This fits with our hypothesis that ST131 have adapted to out-compete commensals for scarce nutrients. It may also explain why previous screening assays did not identify ST131 as metabolically unique, as ST131 had adapted to out-compete commensals rather than utilise more diverse metabolites. The data presented in this chapter has focussed on competition between ST131 and other *E. coli* however the genomic analysis presented in Chapter 3 revealed a metabolic signature that was shared amongst multiple MDR ExPEC lineages. Therefore we would expect other MDR lineages such as ST167 or ST648 to be capable of out-competing commensals in our model. Expanding our analysis to include multiple MDR isolates and competing them against multiple commensal isolates would improve the strength of our observations.

S. Typhimurium is capable of efficiently colonising new hosts and induces the host inflammatory response to provide a new repertoire of metabolites that it is capable of using. We hypothesised that MDR *E. coli* were using the same technique to out-compete commensals and colonise individuals. To test this we examined the inflammatory response from the intestinal epithelial cells. We observed only a minimal response to the presence of ST131 cells and a much stronger IL-8 response to the prescence of K12. Previous work has demonstrated that both cell lines used here (Caco-2 and INT407) are capable of secreting pro-inflammatory cytokines [150, 87, 8]. Previous studies assayed cytokine secretion 24 hours after treatment whilst our data was obtained only after 6 hours [150]. In addition our LPS stimulation also failed to produce a cytokine reponse from the cells however others have demonstrated that INT407 cells secrete a range of cytokines in response to LPS stimulation [8]. Previous studies suggest that combining LPS treatment with exogenous cytokines, specifically IL-1β, can increase the cytokine response from Caco-2 cells [150]. The same study detects a significant increase in IL-8 production from Caco-2 cells when stimulated solely with LPS however this was after 24 hours whilst in our assay cells were only stimulated for 6 hours. Other studies

assessing the inflammatory response of Caco-2 cells to varying strains of bacteria opted to irradiate the bacteria before addition to the cell cultures [87].

Ultimately the *in vitro* models used here suggest that ST131 is capable of out-competing a commensal strain under specific conditions. However, these assays did not address the important contribution that the host inflammatory response makes to pathogen colonisation. Studies examining the colonising ability of *S.* Typhimurium have demonstrated that the host inflammatory response provides a unique source of nutrients which *S.* Typhimurium is genetically equipped to exploit [161, 162, 48]. Importantly, the phenotypic differences conferred by genetic elements only become apparent under specifc conditions [161], specifically under anaerobic *in vitro* conditions, or in mouse models. Here we attempt to model the host response through the use of intestinal epithelial cells however under these conditions an inflammatory cytokine response was not detected. An important component of the host response is the immune system which is responsible for generating a huge diversity of antimicrobial compounds. The immune system is composed of a multitude of cell types which function together to produce an inflammatory response, the influence that these immune cells may exert on bacterial competition is absent in these assays. A more complete model is required to correctly test our hypothesis that MDR lineages of *E. coli* are capable of effectively colonising hosts using the host inflammatory response. Experiments using a murine host model are therefore required.

Chapter 5

Germ-Free Mouse Model of Colonisation Confirms ST131 can Out-compete Commensal *E. coli*

5.1 Introduction

My data in Chapter 3 shows that MDR lineages of *E. coli* display increased variation in metabolic genes, specifically those involved in carbohydrate metabolism and transport as well as an increased abundance of genes in core metabolic pathways. Investigations into the metabolic capabilities of ST131 indicated that this MDR lineage was not metabolically unique [4, 5], however my data in Chapter 4 indicates that ST131 is able to out-compete commensals in *in vitro* assays. These assays were unable to model the effect that the host immune system can have on bacterial growth. Studies of *S.* Typhimurium colonisation identified that it was exploiting the host immune response to outcompete commensals [161]. The host immune system produces reactive oxygen species (ROS) which react with compounds in the gut lumen producing a new variety of metabolites, including tetrathionate. Mice which are unable to produce ROS do not produce this compound and *S.* Typhimurium growth in

these defective mice is impaired [161]. This mechanism is not specific to one compound but has been demonstrated to occur for 1,2-propanediol, ethanolamine and reactive nitrogen species (RNS) [48, 162, 147]. Importantly the genes responsible for RNS utilisation are shared with *E. coli* [85]. Ultimately, the host response is essential for the invading *S.* Typhimurium to effectively colonise the host and can greatly influence the diversity of compounds available to intestinal microbes. To accurately test our hypothesis that MDR *E. coli* out-compete commensal *E. coli* by utilising compounds generated by the host inflammatory response we proceeded to use a mouse model of colonisation.

Attempts to colonise mice with human *E. coli* isolates have been reported to have mixed results, with strains showing a high degree of colonisation variability [79]. Frequently mice are treated with the antibiotic streptomycin in order to permit colonisation by *E. coli*, however this requires the introduction of streptomycin resistance genes to the assayed strains [49]. As we wish to determine whether MDR *E. coli* have an innate advantage over commensal *E. coli* in their ability to colonise hosts the presence of antibiotics, as well as artificially introducing resistance markers, could create a bias favouring MDR *E. coli* which have already adapted to the presence of AMR genes. Moreover treatment with streptomycin has also been observed to alter the host immune response, with treated mice exhibiting increased levels of inflammation [141], which introduces another bias to the model. The authors even suggest that this effect may be contributing to increased levels of colonisation by *E. coli* and *Salmonella*. In addition, previous studies have demonstrated that the introduction of streptomycin resistance genes into *E. coli* alters the expression of virulence factors, again potentially introducing bias into the system [27]. Therefore a germ free mouse model was chosen as it allowed us to examine the colonisation abilities of MDR and commensal *E. coli* in the absence of antibiotics, permitting the use of unaltered strains. We hypothesise that in this model the MDR *E. coli* will be able to out-compete the commensal when introduced into the mouse. Furthermore, we can introduce a commensal strain into the mice first, to mimic the presence of a native

microbiota, and then challenge with an MDR strain. We predict that the MDR *E. coli* will still be able to effectively colonise the host even when a commensal strain is already present. Lastly, we hypothesise that MDR *E. coli* induce a stronger inflammatory response from the host compared to commensal *E. coli*. To test these hypotheses we will use a commensal isolated from a healthy human and two ExPEC isolates one of which will have minimal antibiotic resistance whilst the other will have high levels of resistance. We expect that the ExPEC strain with higher resistance will be the most effective coloniser of the host and will induce the greatest inflammatory response.

5.2 Results

5.2.1 Monocolonisation

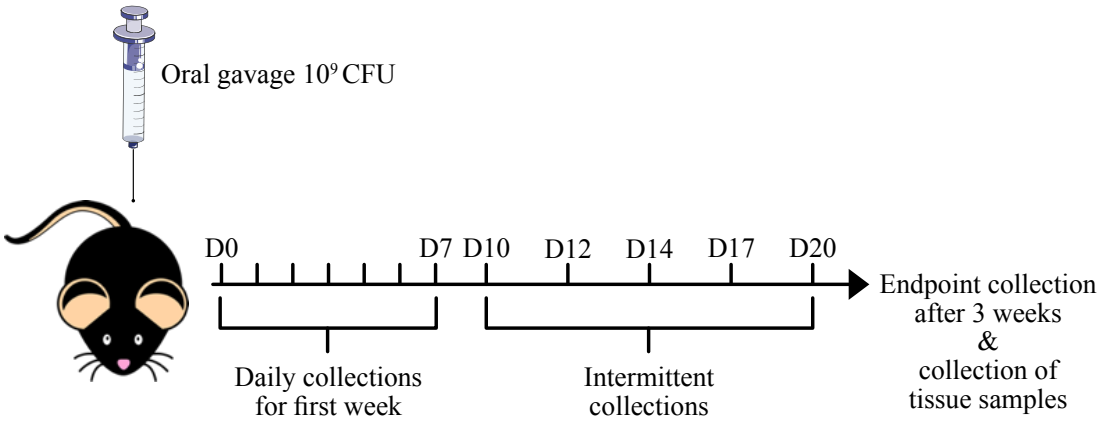

Fig. 5.1 Schematic of the colonisation of mice by a single *E. coli* strain.

Strains are capable of colonising germ-free mice

Germ-free mice were orally gavaged with 1×10^9 CFU of bacteria and colonisation was monitored by enumerating CFU in the faecal pellets for 3 weeks following colonisation (Figure 5.1). Three different *E. coli* strains were assayed; a commensal isolated from a healthy

human (811E7, ST10), a clinical isolate from a bacteraemia patient that had minimal antibiotic resistance (F084, ST73) and a clinical isolate from a bacteraemia patient with a high level of resistance (F016, ST131). Mice were successfully colonised by all three strains as observed by bacterial growth in the faecal pellets (Figure 5.2). The CFU in faecal pellets rose over the first 4 days of colonisation before gradually decreasing over the subsequent 5 days to reach stable levels. The mice remained colonised for the duration of the 3 week period.

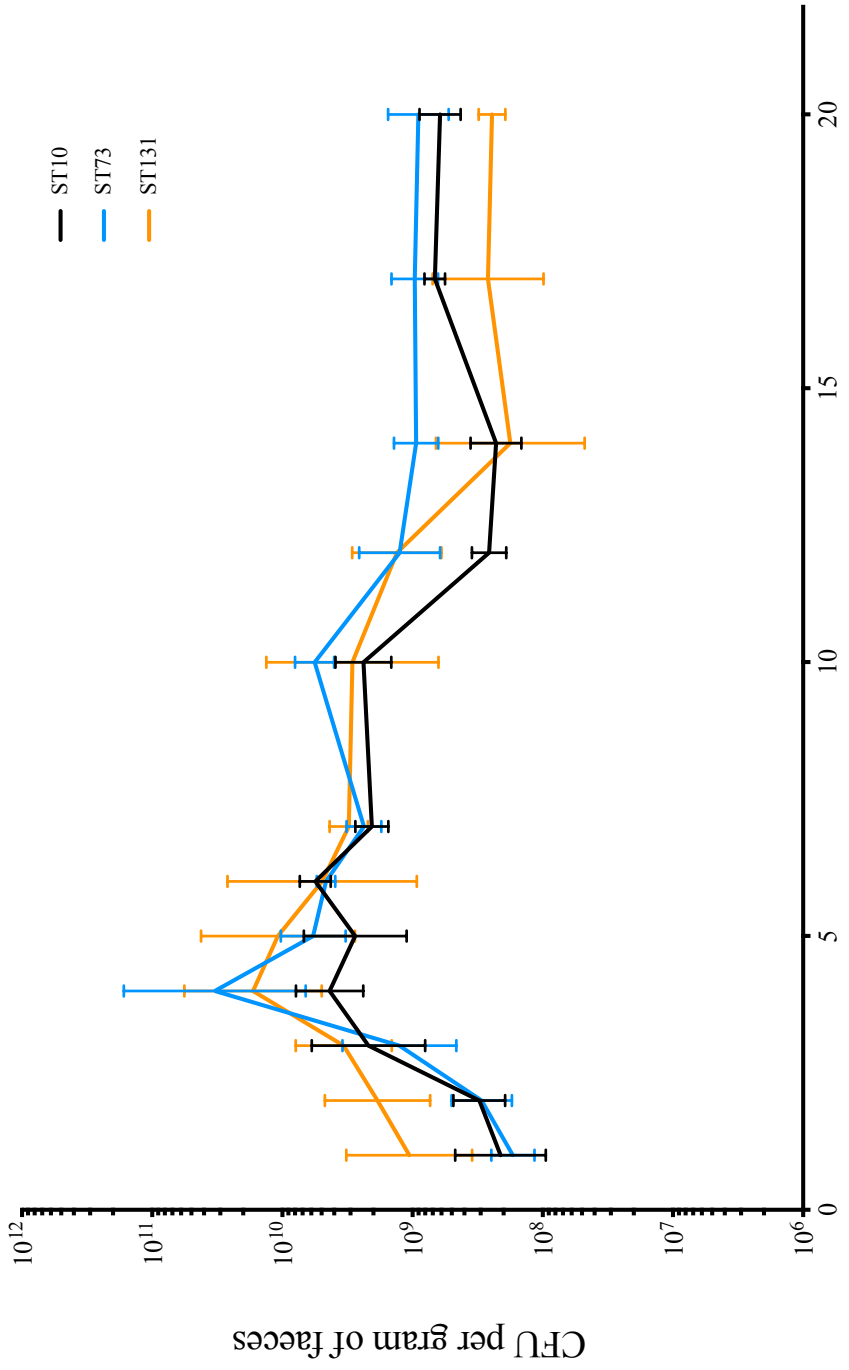

Fig. 5.2 Levels of CFU in the faecal pellets of monocolonised mice over a 3 week period. Lines indicate geometric mean of CFU values, error bars indicate SD. Lines and bars are coloured by strain with black for ST10, blue for ST73 and orange for ST131. n=3

Strains predominantly colonise the caecum and colon, and display only minimal dissemination

After 3 weeks of colonisation the mice were sacrificed for tissue collection. The contents of the small intestine, caecum and colon were collected and assayed for levels of colonisation. The highest CFU values were recorded in the caecum with the commensal ST10 achieving the highest CFU (1.37×10^9 CFU per gram) followed by ST73 (1.13×10^9 CFU per gram) then ST131 (4.54×10^8 CFU per gram) (Figure 5.3a). The colon showed lower levels of colonisation with the commensal ST10 being the most abundant (6.26×10^8 CFU per gram) followed by ST73 (4.96×10^8 CFU per gram) then ST131 (1.01×10^8 CFU per gram). Finally the small intestine had the lowest levels of colonisation with the ST10 the most abundant (2.35×10^7 CFU per gram) followed by ST73 (1.24×10^7 CFU per gram) then ST131 (6.10×10^6 CFU per gram). Alongside the gut contents multiple tissues were also harvested and homogenised to determine if any of the strains had disseminated to extra-intestinal sites. Surprisingly the commensal ST10 strain showed the highest levels of dissemination with multiple mice having detectable bacteria present in their spleen, Mesenteric Lymph Nodess (MLNs), lung, brain and blood (Figure 5.3b). None of the mice showed signs of disease during the experiment. The ST73 and ST131 strains used did not exhibit the same levels of dissemination with only two ST73 and two ST131 colonised mice showing dissemination to the MLN and one mouse from each of the ST73 and ST131 groups showing colonisation of the lung. Again, only one ST73 colonised mouse had detectable bacteria in the blood. The surprising observation that the commensal ST10 was detectable in multiple extra-intestinal tissues could be explained by contaminated surgical instruments. Perforating the intestine with surgical instruments during tissue harvesting can easily spread bacteria to otherwise uninfected tissues.

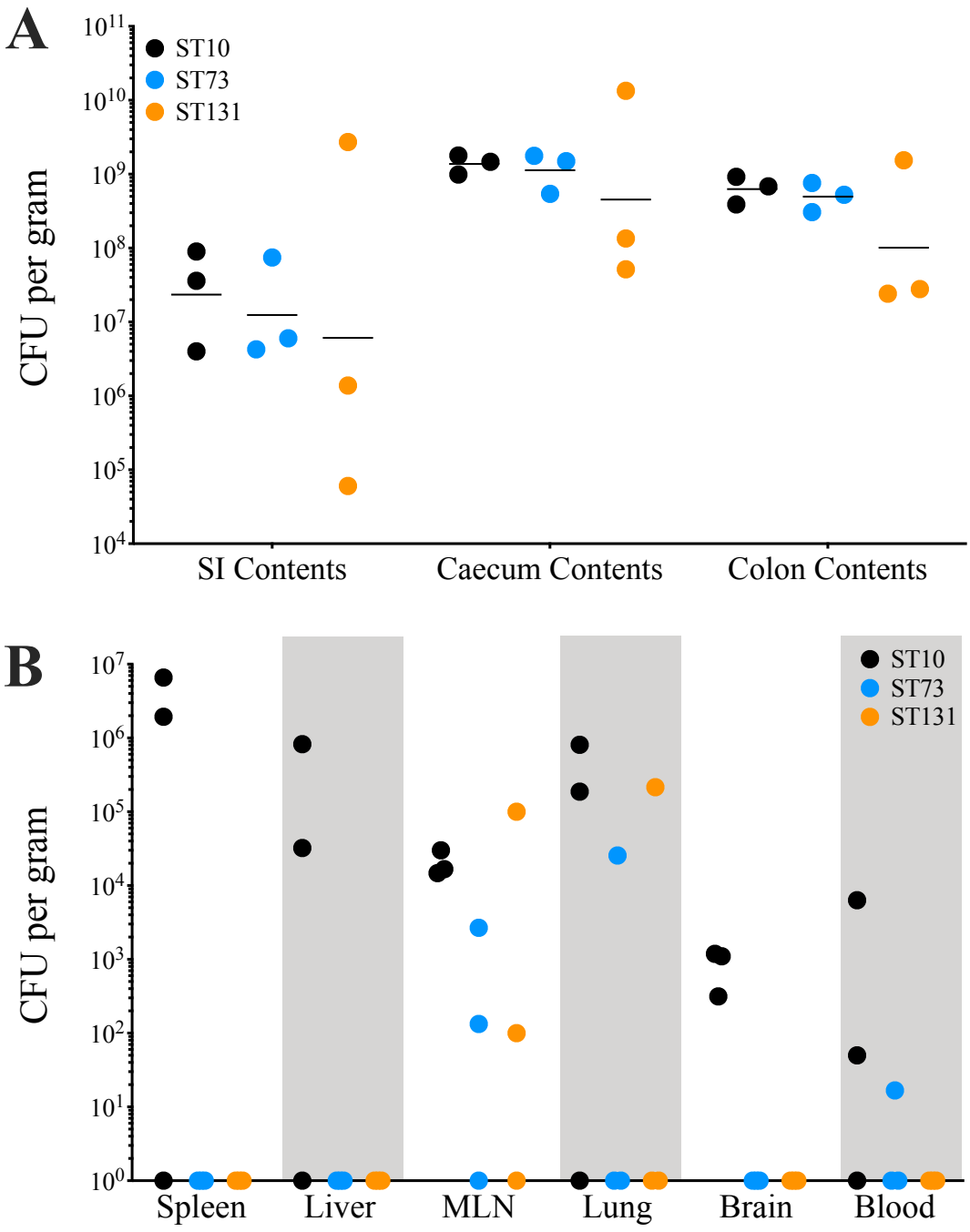

Fig. 5.3 CFU Values from Gut Contents and Extra-Intestinal Tissues. Points indicate values for individual mice and are coloured according to colonising strain (ST10: black, ST73: blue, ST131: orange). The CFU per gram of gut contents from the small intestine (SI), caecum and colon (A). Lines indicate geometric mean. The CFU per gram of tissue for multiple extra-intestinal tissues (B). n=3

Colonised mice have altered cytokine expression after 3 weeks

Sections of the small intestine, caecum and colon were collected and total RNA was isolated. Expression of multiple pro- and anti-inflammatory cytokines as well as the S100A8 subunit of the antimicrobial peptide were assayed using probe-based qPCR. The mRNA levels of 11 assayed genes were normalised to the expression of the housekeeping gene *Pol2ra* (which encodes an RNA polymerase subunit). Mice colonised with the commensal ST10 showed significantly increased levels of the pro-inflammatory cytokines *IL-1α*, *IL-17* and *IL-33* when compared to germ-free control mice, the expression of *S100A8* is also significantly increased (Figure 5.4a). IL-1α and IL-33 both function as alarmins which are released from damaged epithelial cells [13, 14], while IL-17 is produced by several immune cell types and functions to promote an inflammatory response from multiple cells [167]. Colonisation with ST73 did not affect cytokine expression in the small intestine, nor did ST131. We did not expect the commensal strain to produce the strongest inflammatory response in the small intestine.

In the caecum, *IL-17* expression was significantly increased in mice colonised with ST131 compared to both ST10 and ST73 colonised mice, as well as germ-free mice (Figure 5.4b). Commensally colonised mice had a significant increase in *IL-23a* expression in the caecum compared to germ-free mice. IL-23a is produced from Dendritic Cells (DCs) and macrophages and acts upon T cells, Natural Killer (NK) cells and Innate Lymphoid Cells (ILCs) to promote the expression of IL-17 [167], however no difference in *IL-17* expression is observed in these mice. In the colon of ST73 colonised mice expression of *IL-17* was significantly increased compared to commensally colonised and germ-free mice, while expression of the other cytokines was unaffected (Figure 5.4c). To further examine the extent of the host response tissues sections from the small intestine and colon were formaldehyde fixed, embedded in paraffin and sectioned. Tissue sections were stained with haematoxylin and eosin (H&E) before imaging to examine markers of inflammation such as infiltration of

inflammatory cells, changes to the intestinal epithelium and overall alterations to mucosal architecture. Comparing these sections revealed there was no difference between colonising strains (Figure 5.5). Blind-scoring the images using an inflammatory scoring system for IBD and colitis confirmed there was no difference between colonisation conditions [45] (Figure 5.6).

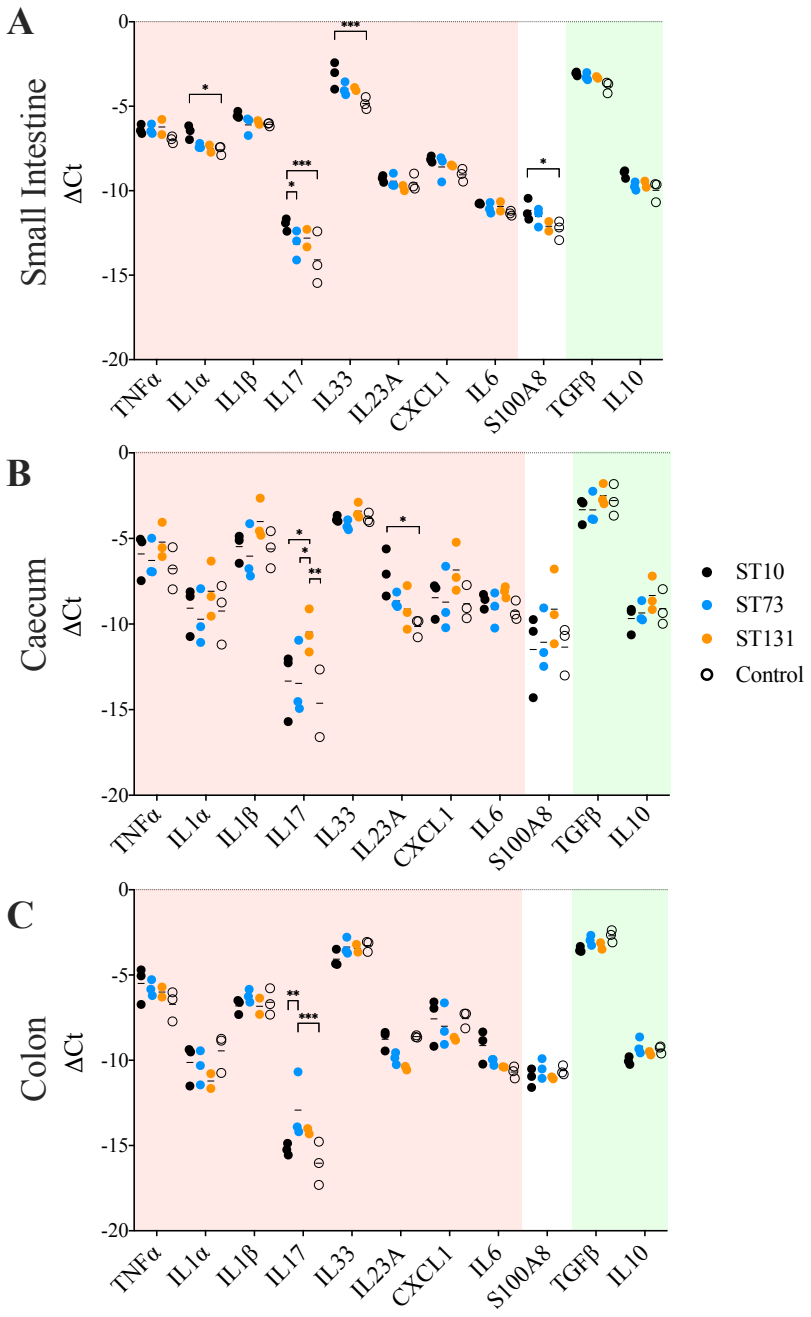

Fig. 5.4 Expression of multiple cytokines assayed by probe based qPCR. Ct values are normalised to the housekeeping gene *Pol2ra* (ΔCt). Points indicate values for individual mice and are coloured according to colonisation (ST10: black, ST73: blue, ST131: orange, germ free: open circles). Lines indicate mean. Cytokines are clustered into pro-inflammatory (red box) and anti-inflammatory (green box). Sections of the small intestine (A), caecum (B) and colon (C) were assayed. Significance determined by 2-way ANOVA with Tukey's multiple comparison, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. n=3

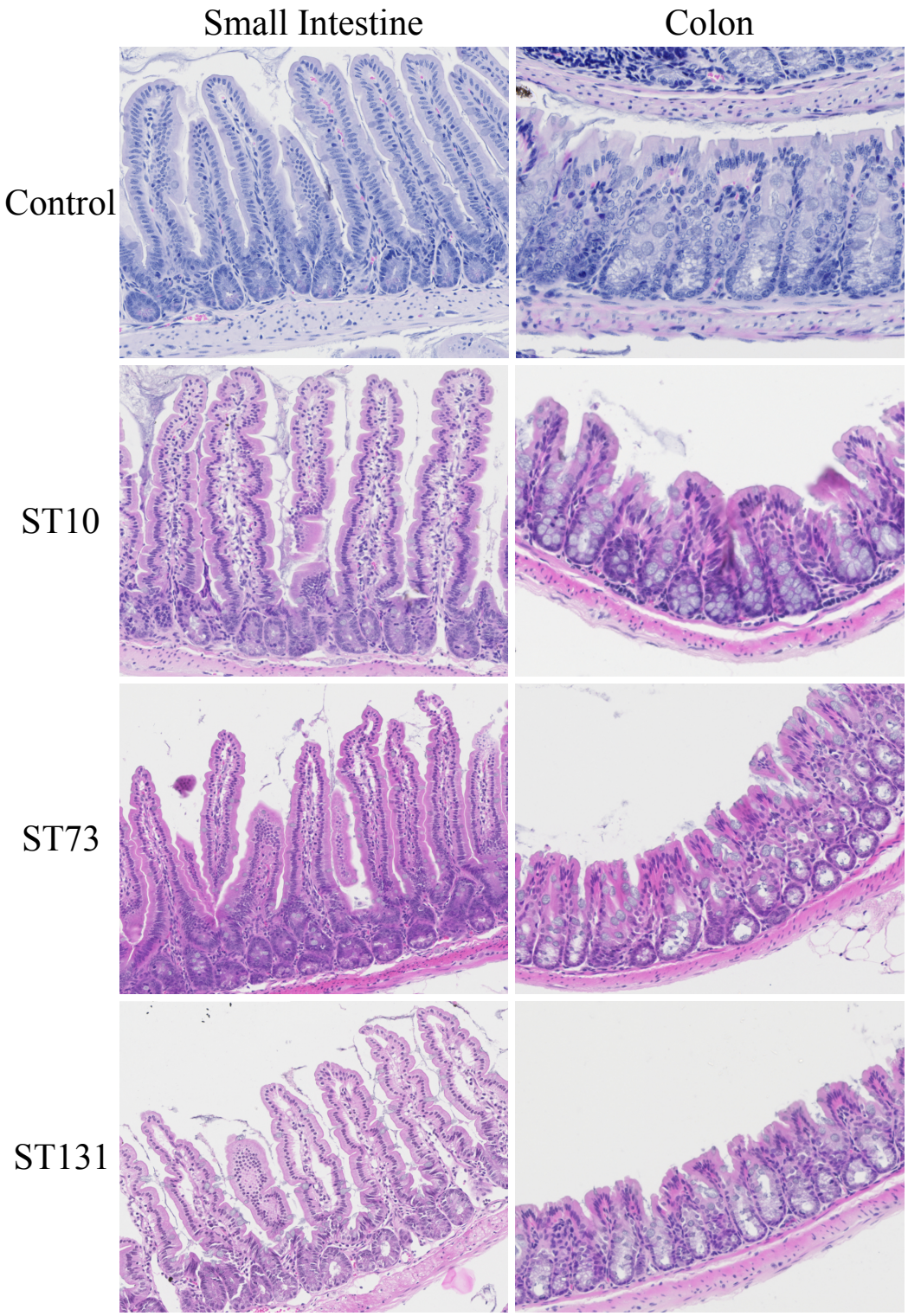

Fig. 5.5 Representative Tissue Sections of Small Intestine and Colon from Mono-Colonised Mice. Tissue sections have been H&E stained.

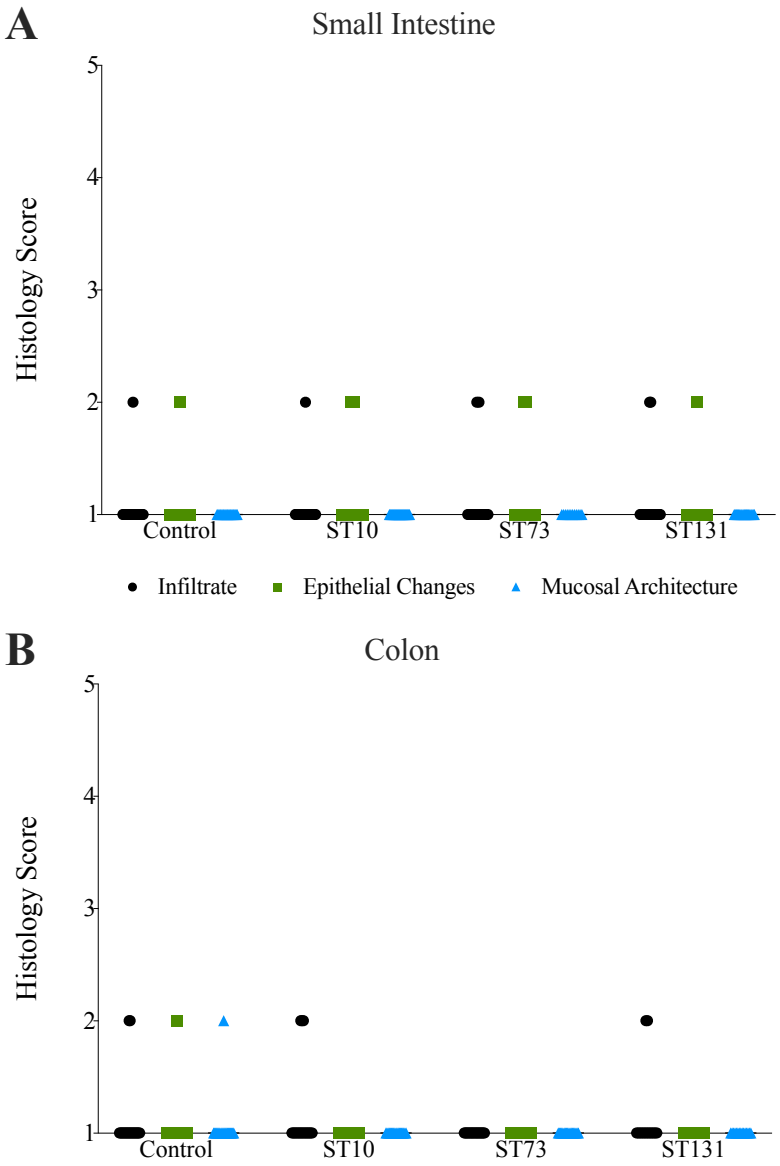

Fig. 5.6 Tissue Pathology Scores from Mono-Colonised Mice. Sections of small intestine (A) and colon (B) were blind scored using a scoring system for IBD mouse models. Scores range from 1 (absent / minor) to 5 (severe) for the categories: inflammatory infiltrate (black circles), changes to epithelium (green square) and alterations to overall mucosal architecture (blue triangles).

5.2.2 Competitive Colonisation

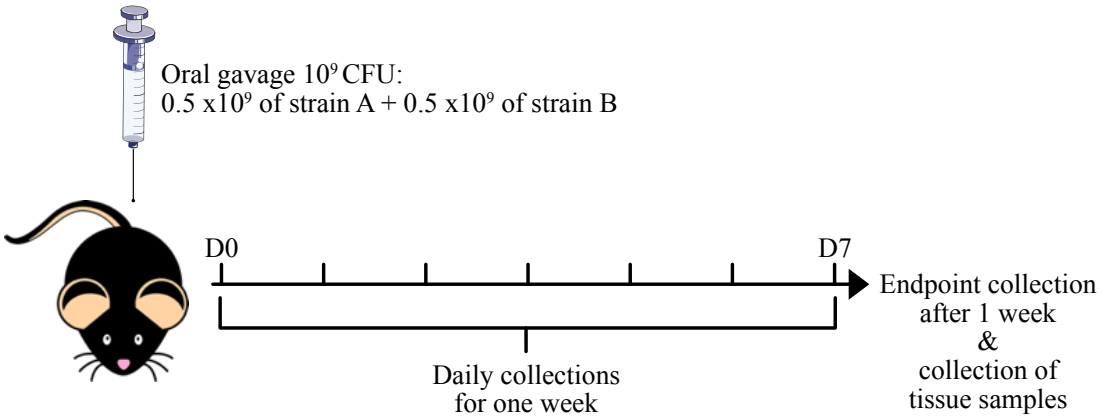

Fig. 5.7 Schematic of the colonisation of mice by two *E. coli* strains simultaeneously.

ST131 out-competes commensals to colonise germ-free mice

Mono-colonisation experiments demonstrated that all the strains assayed were capable of colonising germ-free mice and did so to the same extent. To test our hypothesis that MDR strains are better able to colonise hosts than commensal strains, we competed an MDR strain against a commensal. Strains were mixed in a 1 to 1 ratio immediately prior to gavaging the mouse (Figure 5.7). Again colonisation was monitored by enumerating the CFU present in faecal pellets; strains were differentiated using selective agar plates. As observed previously all strains were capable of colonising the mice with CFU levels increasing over the first week of colonisation (Figure 5.8). When the commensal ST10 was competed against the ST73 both strains were able to colonise the mouse. The ST73 strain accounted for greater than 50% of the total growth for the first 5 days before rising to account for close to 100% of growth by day 6 (Figure 5.8a). Demonstrating that the ST73 strain has a slight advantage over the commensal. When the ST10 and ST131 are co-gavaged into the mouse the ST131 strain achieves near 100% dominance over the commensal by day 4 indicating that it has a stronger competitive advantage over the commensal than the ST73 strain (Figure 5.8b).

Lastly, the ST73 and ST131 strains were competed against one another, under these conditions the ST131 strain accounted for slightly more than 50% of the growth but this did not change over the course of the experiment demonstrating that ST131 has a slight advantage over ST73 (Figure 5.8c).

As observed in previous experiments, levels of colonisation were higher in the caecum than the colon. Levels of colonisation were similar in the caecum between all conditions assayed (Figure 5.9). When the commensal ST10 was competed against the ST73 strain there was an average total growth of 1.93×10^{10} CFU per gram of which 1.45×10^{10} CFU per gram was ST73. Competing ST10 against ST131 produced a total growth of 1.59×10^{10} CFU per gram of which 1.46×10^{10} CFU per gram was from ST131. While competing ST73 against ST131 produced a total growth of 2.25×10^{10} CFU per gram of which 1.43×10^{10} CFU per gram was from ST131. Levels of bacteria in the colon were lower, consistent with previous experiments (Figure 5.9a). Co-gavaging ST10 and ST73 resulted in 3.60×10^{9} CFU per gram in the colon of which 2.95×10^{9} CFU per gram was from ST73. Co-gavaging ST10 and ST131 produced a total growth of 4.14×10^{9} CFU per gram of which 3.06×10^{9} CFU per gram was ST131. Finally co-gavaging ST73 and ST131 produced a total growth of 5.78×10^{9} CFU per gram of which 4.48×10^{9} CFU per gram was from ST131. Examining this data as an average percentage shows that ST73 was able to out-compete the commensal as it was responsible for 84.7% of growth from the caecum but the ST131 was able to achieve 94.1% demonstrating it is better at out-competing a commensal (Figure 5.9b). ST131 is only able to slightly out-compete the ST73 strain accounting for 64.2% of growth on average. In the colon the ST73 and ST131 performed at equitable levels when competed against a commensal with the ST73 accounting for an average of 84.7% of total growth and ST131 accounting for 78.4%. When ST73 and ST131 were competed against one another the ST131 out-grew the ST73 in the colon as it accounted for an average of 81.4% of growth.

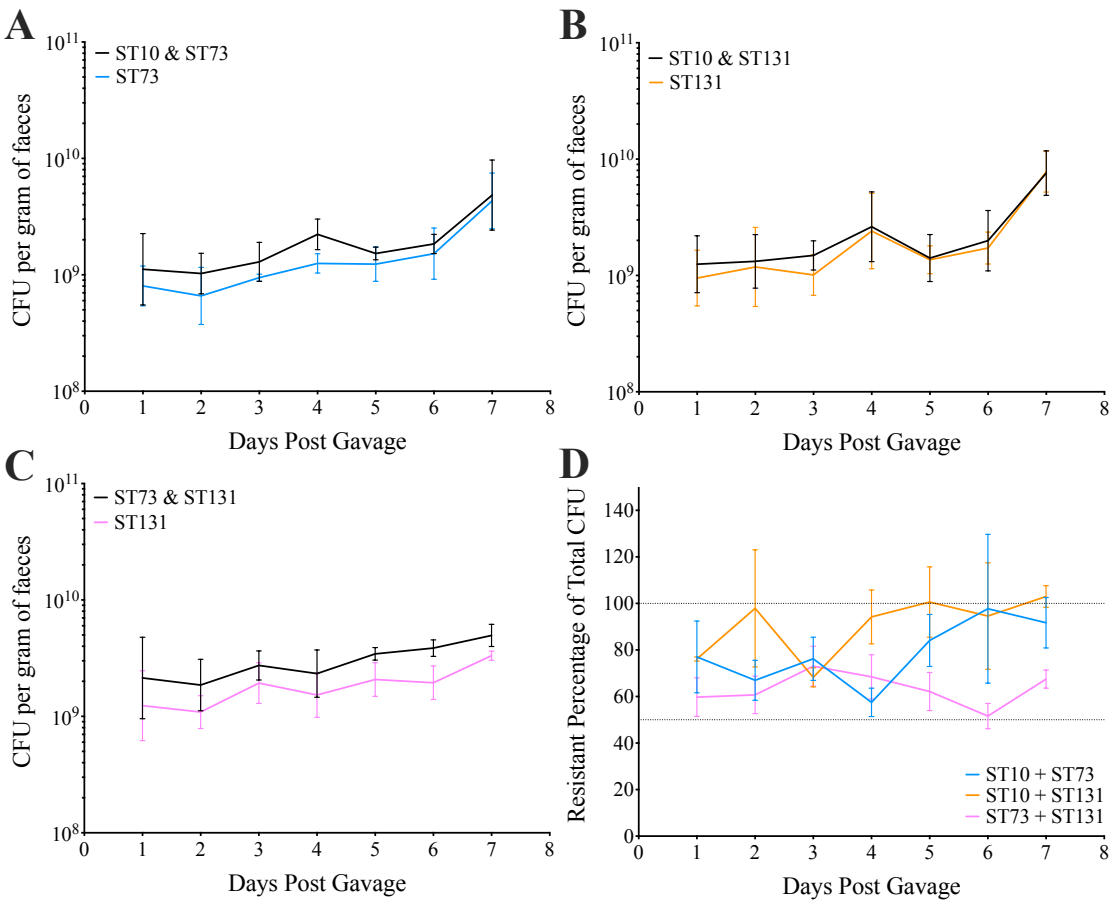

Fig. 5.8 Levels of CFU in the faecael pellets of co-gavaged mice over a 1 week period. Lines indicate geometric mean, error bars indicate SD. A) Mice that were co-gavaged with ST10 and ST73. Black line indicates total growth of both strains (non-selective agar), blue line indicates growth of solely ST73 (selective agar). B) Mice that were co-gavaged with ST10 and ST131. Black line indicates total growth, orange line indicates growth of ST131. C) Mice that were co-gavaged with ST73 and ST131. Black line indicates total growth, pink line indicates growth of ST131. D) Summarisation of panels A, B and C. Percentage of growth that is attributable to resistant strain, calculated by dividing resistant CFU growth by total CFU growth. Blue line indicates ST10 & ST73 mice, orange indicates ST10 & ST131 mice and pink indicates ST73 & ST131 mice.

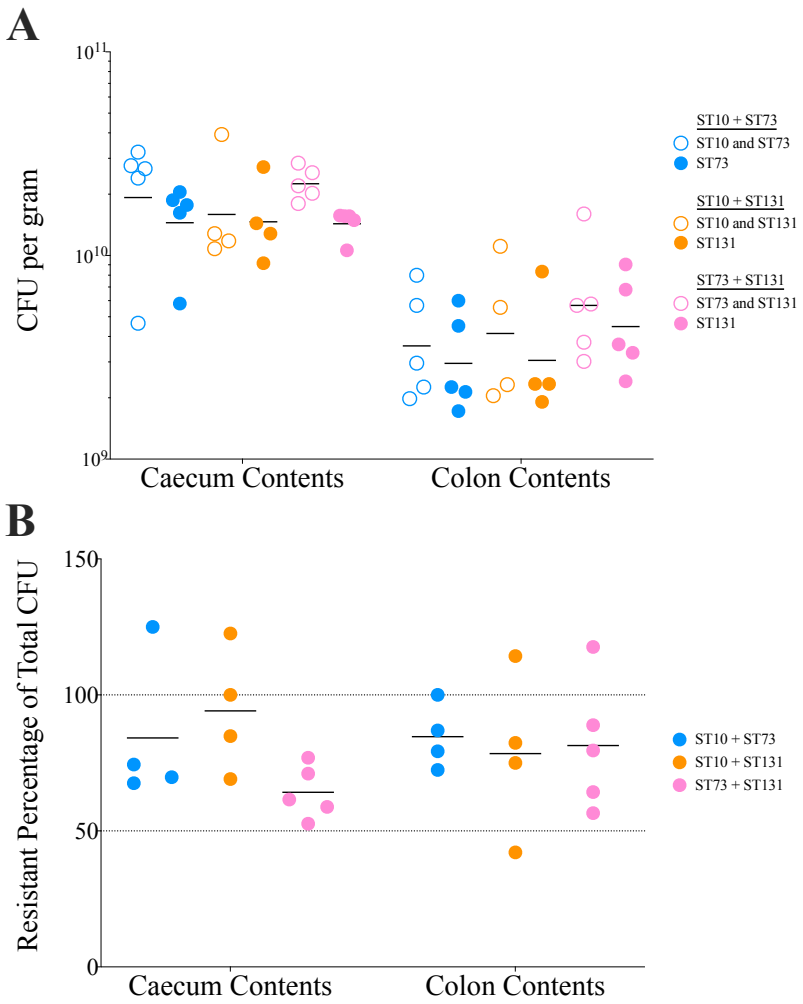

Fig. 5.9 CFU Values from Gut Contents of Co-gavaged Mice. Points represent values from individual mice and are coloured by colonisation condition (ST10 & ST773: blue, ST10 & ST131: orange, ST73 & ST131: pink). Open circles indicate total growth of both strains while closed circles indicate growth on selective agar. A) Raw CFU values per gram of gut contents. Lines indicate geometric mean. B) Percentage of total growth that is attributable to resistant strain, calculated by dividing growth on selective agar by total growth on non-selective agar. Lines indicate mean value.

Low levels of strain dissemination after 1 week

Dissemination of bacteria to extra-intestinal sites was measured as before however only the primary dissemination sites were assayed (MLN, spleen, liver and blood). Overall there was minimal dissemination of bacteria to the spleen, liver, MLN and blood compared to the monocolonised mice (Figure 5.10). None of the mice gavaged with a mixture of commensal ST10 and ST73 had detectable bacteria in extra-intestinal sites (Figure 5.10a). Only one mouse gavaged with the ST10 and ST131 mixture had detectable bacteria in the MLN, the majority of which was ST131 (Figure 5.10b). One mouse which was gavaged with the ST73 and ST131 mixture had dissemination of both strains to the spleen (Figure 5.10c). This indicates that one week after colonisation strains rarely colonise extra intestinal tissues.

Colonised mice show altered cytokine response by qPCR

Differences in cytokine expression were examined by qPCR. All colonisation conditions assayed showed significantly increased expression of *IL-17* in the small intestine compared to germ-free mice (Figure 5.11a). There was a significant increase in the expression of *S100A8* in the small intestine of mice colonised by ST73. Analysing the cytokine expression in the caecum of these mice showed that colonisation, by any strain, significantly altered the expression of the pro-inflammatory cytokines *TNFα*, *IL-1α*, *IL-1β*, *IL-17*, *CXCL1* and *IL-6* as well as *S100A8* when compared to germ-free mice (Figure 5.11b). Additionally *IL-10* was significantly higher in mice colonised with ST10 and ST131 compared to germ free mice but was not altered under other colonisation conditions. IL-10 is produced by regulatory T cells which suppress inflammation and its presence is associated with tolerance [110]. There were no significant differences between colonisation by different strains. In the colon there was significant alteration of *TNFα*, *IL-1β*, *IL-17*, *CXCL1* and *S100A8* expression in colonised mice versus germ-free mice (Figure 5.11c). The expression of *IL-6* was also significantly altered in mice colonised with ST73 and ST131 versus germ-free mice. Again there were

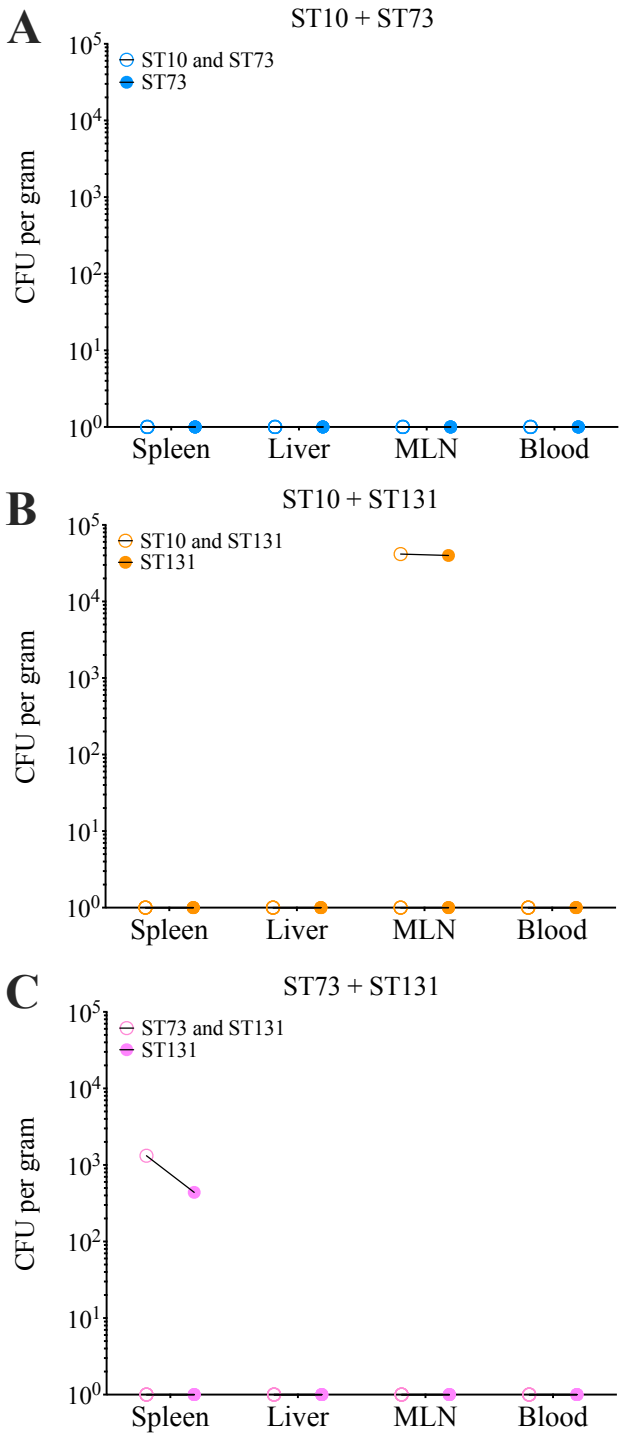

Fig. 5.10 CFU Values from Extra-Intestinal Tissues from Co-gavaged Mice. Points represent CFU values from individual mice and are coloured by colonisation condition with blue representing ST10 & ST73 (A), orange representing ST10 & ST131 (B) and pink representing ST73 & ST131 (C). Open circles indicate total growth of both strains while closed circles indicate growth on selective agar. Lines connect CFU measurements from the same mouse.

no observed differences between colonisation by different strains. This data suggests that the mice have a pro-inflammatory response to the presence of ExPEC strains regardless of whether a commensal is also present. Moreover, ST131 does not elicit a stronger inflammatory response than ST73 .

Sections of the small intestine and caecum were examined for histological markers of inflammation. No difference was observed between the colonisation conditions nor between germ-free mice (Figure 5.12). The sections were scored as before but no difference was detected (Figure 5.13).

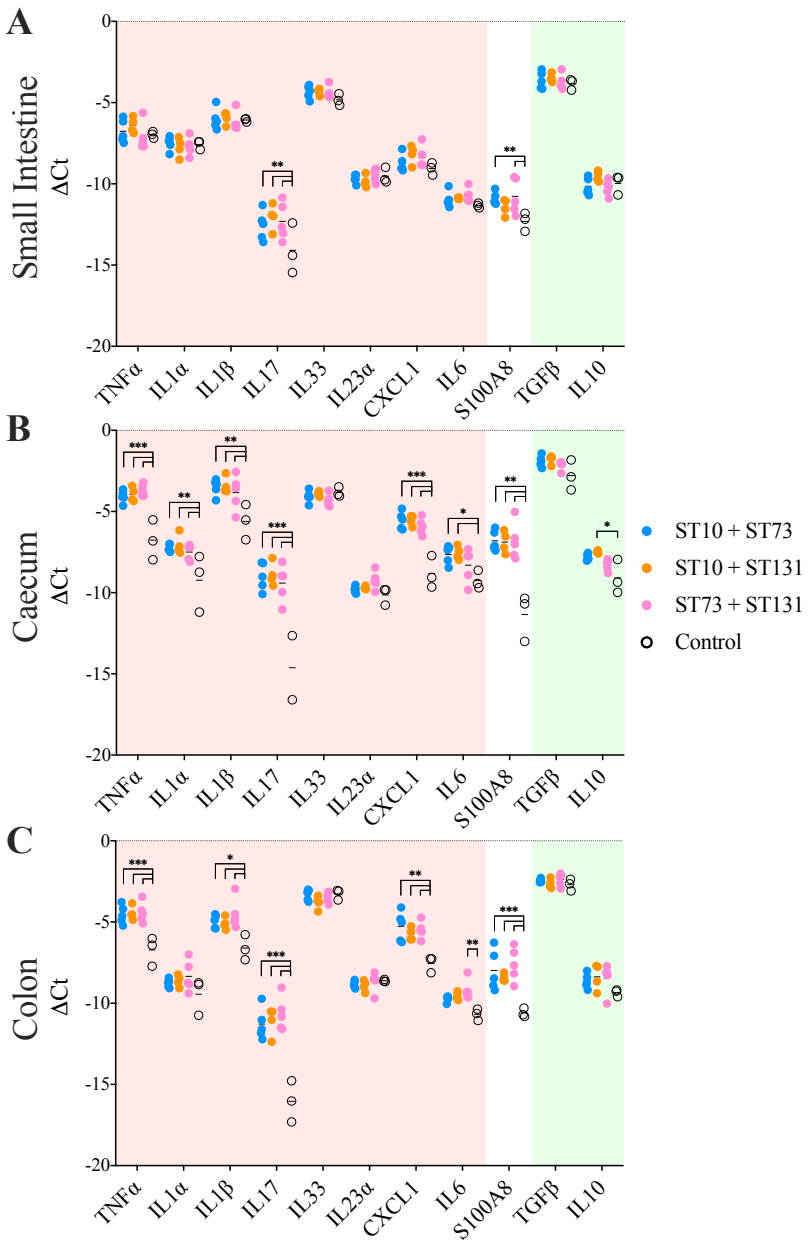

Fig. 5.11 Expression of multiple cytokines as assayed by probe based qPCR. Ct values are normalised to the housekeeping gene *Pol2ra* (ΔCt). Points indicate values for individual mice and are coloured according to colonisation (ST10 & ST73: blue, ST10 & ST131: orange, ST73 & ST131: pink, germ free: open circles). Lines indicate mean value. Cytokines are grouped into pro-inflammatory (red box) and anti-inflammatory (green box). Sections of the small intestine (A), caecum (B) and colon (C) were assayed. Significance determined by 2way ANOVA with Tukey's multiple comparison, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. n=5

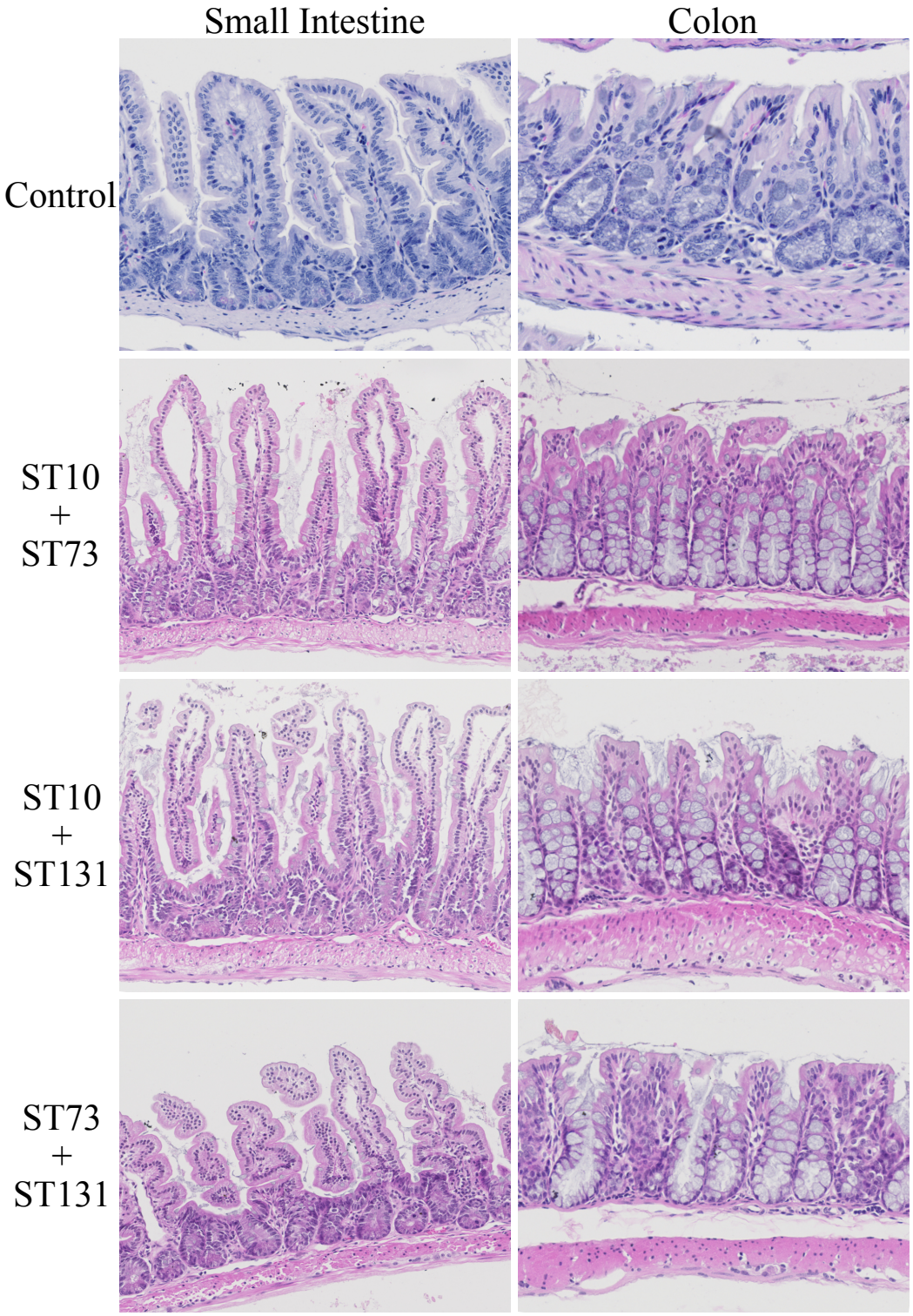

Fig. 5.12 Representative Tissue Sections of Small Intestine and Colon from Co-gavaged Mice. Tissue sections have been H&E stained. Control group is non-colonised germ free mice.

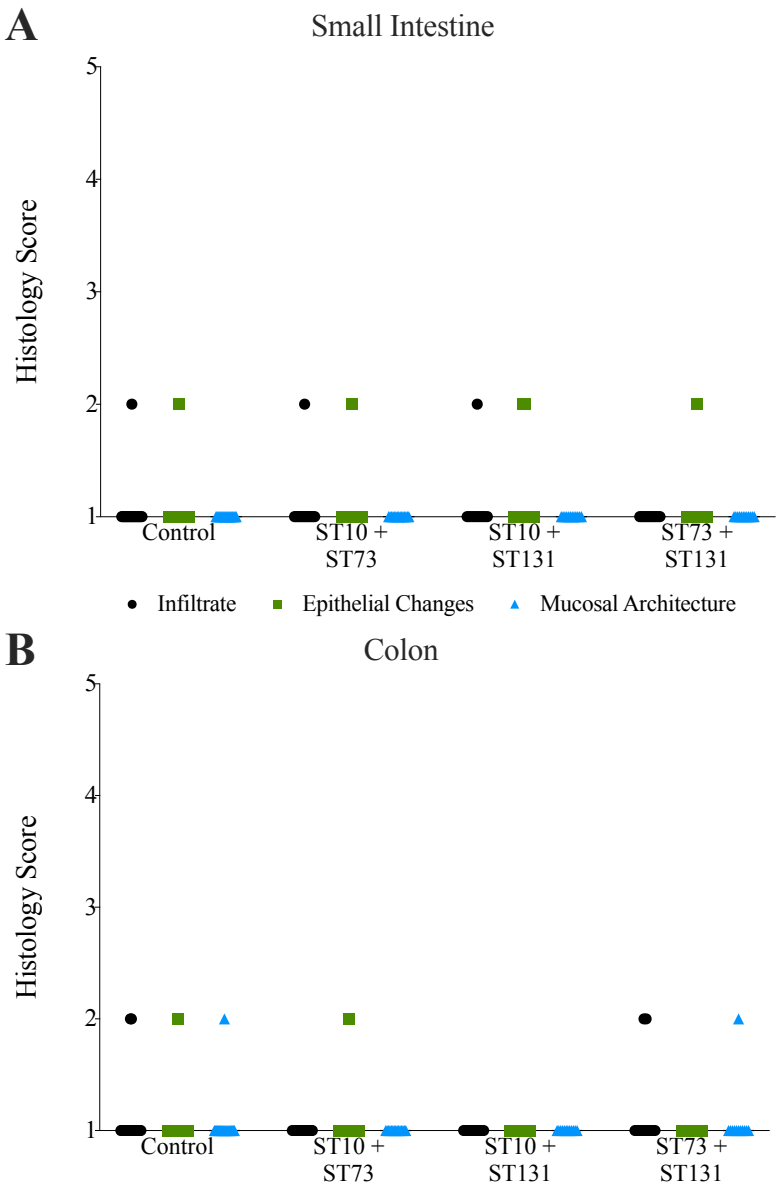

Fig. 5.13 Tissue Pathology Scores from Co-gavaged Mice. Sections of small intestine (A) and colon (B) were blind scored using a scoring system for IBD mouse models. Scores range from 1 (absent / minor) to 5 (severe) for the categories: inflammatory infiltrate (black circles), changes to epithelium (green square) and alterations to overall mucosal architecture (blue triangles)

5.2.3 Displacement of Resident *E. coli*

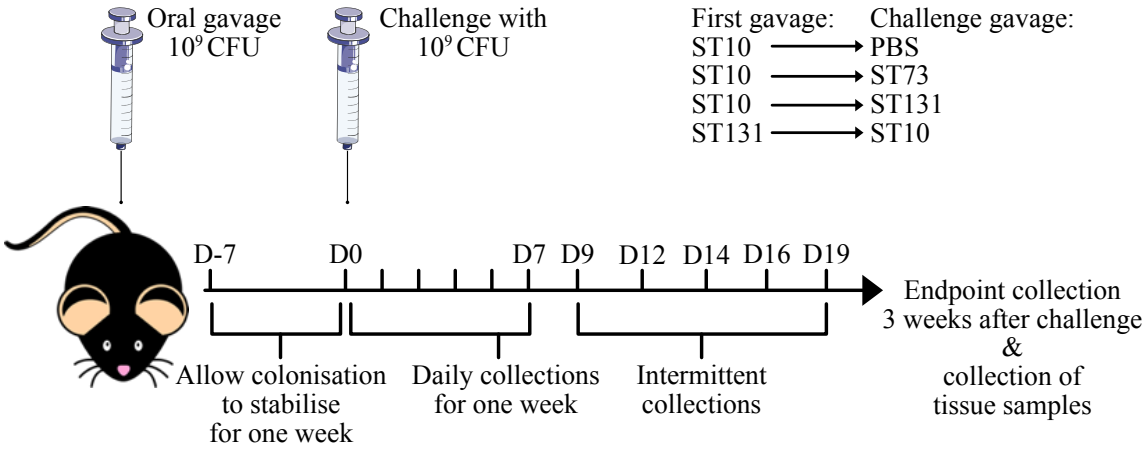

Fig. 5.14 Schematic of the colonisation of mice by one *E. coli* strain followed by challenge with a second strain after 1 week.

ST131 is able to displace a resident commensal ST10

These data indicate that ST131 is able to colonise germ-free mice better than a commensal strain when the two are gavaged at the same time. However, the ST131 strain would normally be an invading bacterium that must compete with an already established commensal. To test whether ST131 was capable of this, mice were colonised with the commensal ST10 and the colonisation was allowed to stabilise for 1 week before challenging with a second strain (Figure 5.14).

Under these conditions, when ST131 was introduced into an ST10 colonised mouse the ST131 strain rose to account for near 100% of total growth by day 3 post challenge, consistent with the colonisation dynamics from the co-gavaged mice (Figure 5.15). ST131 continued to account for 100% of the total growth till day 10 post challenge after which ST131 growth started to decline (Figure 5.15e). By day 21 post challenge ST131 accounted for less than 50% of total growth. The raw CFU counts also support this observation as the mice that were challenged with ST131 show an increase in CFU from 5.36×10^8 CFU/g to 2.06×10^9

CFU/g immediately following challenge (Figure 5.15c). This higher level of colonisation was maintained during the period in which ST131 accounted for 100% of total growth. As the percentage of growth attributable to ST131 fell in these mice, so did the total CFU measured in the faeces. When ST73 was introduced into an ST10 colonised mouse, the invading ST73 strain was also able to dominate over the resident commensal accounting for greater than 50% of the total growth for the first 6 days post challenge, but failed to consistently account for 100% of the total growth (Figure 5.15e). Moreover, raw CFU values indicate that when ST73 was used to challenge the commensal strain the total growth was only altered for the first day post challenge after which levels of colonisation remained stable till the end point (Figure 5.15b). As a control group, mice were first colonised with ST131 for one week before challenging with ST10. In this group the invading ST10 was unable to displace the resident ST131 instead the two strains coexisted with ST131 accounting for 50% of total growth (Figure 5.15e). This demonstrates that the introduction of a large number of invading bacteria is not responsible for the observed dominance of ST131 over the commensal. As an additional control, mice were gavaged with PBS to determine whether the gavage process adversely altered commensal colonisation. No alteration of colonisation was observed in this group thereby eliminating the possibility that the procedure affected commensal growth (Figure 5.15a). Therefore, when ST131 is introduced into a pre-colonised mouse it is able to displace the resident ST10 strain for 10 days before subsiding, however, ST131 is not displaced by an invading ST10 strain.

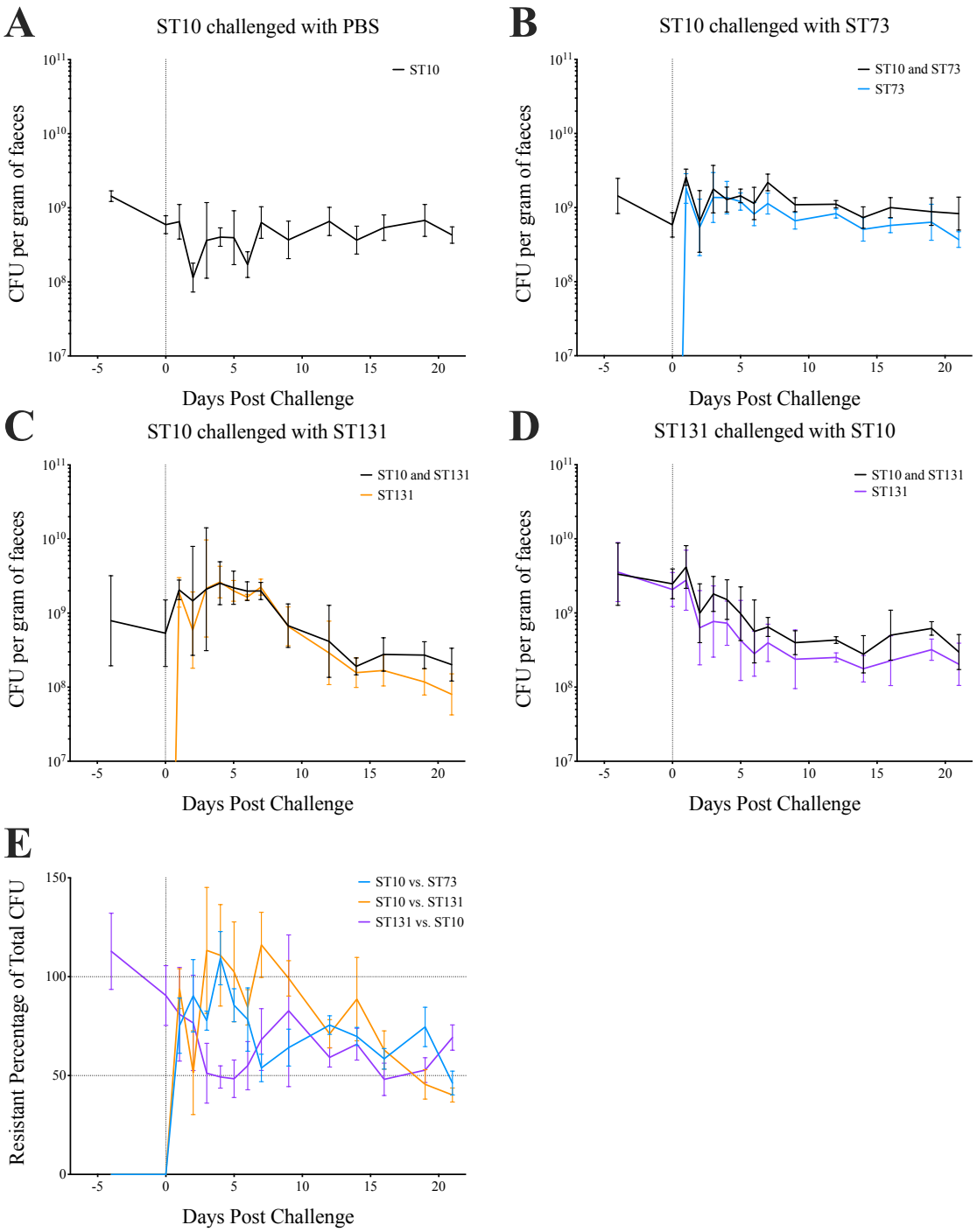

Fig. 5.15 Levels of CFU in the faecael pellets of challenged mice. Lines indicate geometric mean, error bars indicate SD. Black lines indicates total growth of both strains (non-selective agar). Coloured lines indicate growth of ExPEC strain. Mice were challenged with second strain on Day 0. A) Mice that were colonised with ST10 then challenged with PBS. B) Mice that were colonised with ST10 then challenged with ST73 . C) Mice that were colonised with ST10 then challenged with ST131. D) Mice that were colonised with ST131 then challenged with ST10. E) Summarisation of panels B, C and D as percentage of growth attributable to resistant strain.

Commensal out-grows ST131 in the caecum but growth remains equitable in the colon

Levels of colonisation remain higher in the caecum than the colon in the challenged mice consistent with our previous observations (Figure 5.16a). All colonisation conditions display a similar amount of total growth from the caecal contents. There was a total growth of 2.46×10^9 CFU/g for the mice challenged with PBS of which there was no detectable growth on selective agar. For the ST73 challenged mice there was a total growth of 3.99×10^9 CFU/g of which 2.04×10^9 CFU/g was from ST73. There was a total growth of 2.60×10^9 CFU/g from the ST131 challenged mice of which 5.99×10^8 CFU/g was from ST131. Lastly, the ST10 challenged mice produced a total growth of 2.65×10^9 CFU/g of which 1.17×10^9 CFU/g was ST131. On average ST73 accounted for 52.6% of the growth in the caecum while ST131 only accounted for 23.4% (Figure 5.16b). For the control ST10 challenged mice ST131 accounted for 46.5% of growth. The drop in ST131 levels in the caecum indicate that it is struggling to compete with the commensal 3 weeks after invading. Curiously when ST10 was the invading bacterium ST131 was able to persist after 3 weeks suggesting the two strains have a different relationship if ST131 is already established in the host. Alternatively the host may be responding differently to an invading ST131 bacterium compared to a resident ST131.

Total growth recovered from the colon was lower than in the caecum, which is consistent with previous observations (Figure 5.16a). There is an average total growth of 6.50×10^8 CFU/g for the PBS challenged mice with no detectable growth on selective agar. The ST73 challenged mice produced a total growth of 5.24×10^8 CFU/g of which 3.46×10^8 CFU/g was from ST73. In the ST131 challenged mice there was a total growth of 4.40×10^8 CFU/g of which 2.54×10^8 CFU/g was attributable to ST131. Lastly, the ST10 challenged mice produced a total growth of 3.95×10^8 CFU/g of which 2.68×10^8 CFU/g was from ST131. Therefore, ST73 accounted for 70.0% of the growth on average in the ST73 challenged mice (Figure 5.16b). Despite the drop in ST131 growth in the caecum ST131 growth in the colon

was equitable to the commensal with an average of 58.9% of total growth. Lastly, 68.6% of the growth observed in the ST10 challenged mice was from ST131.

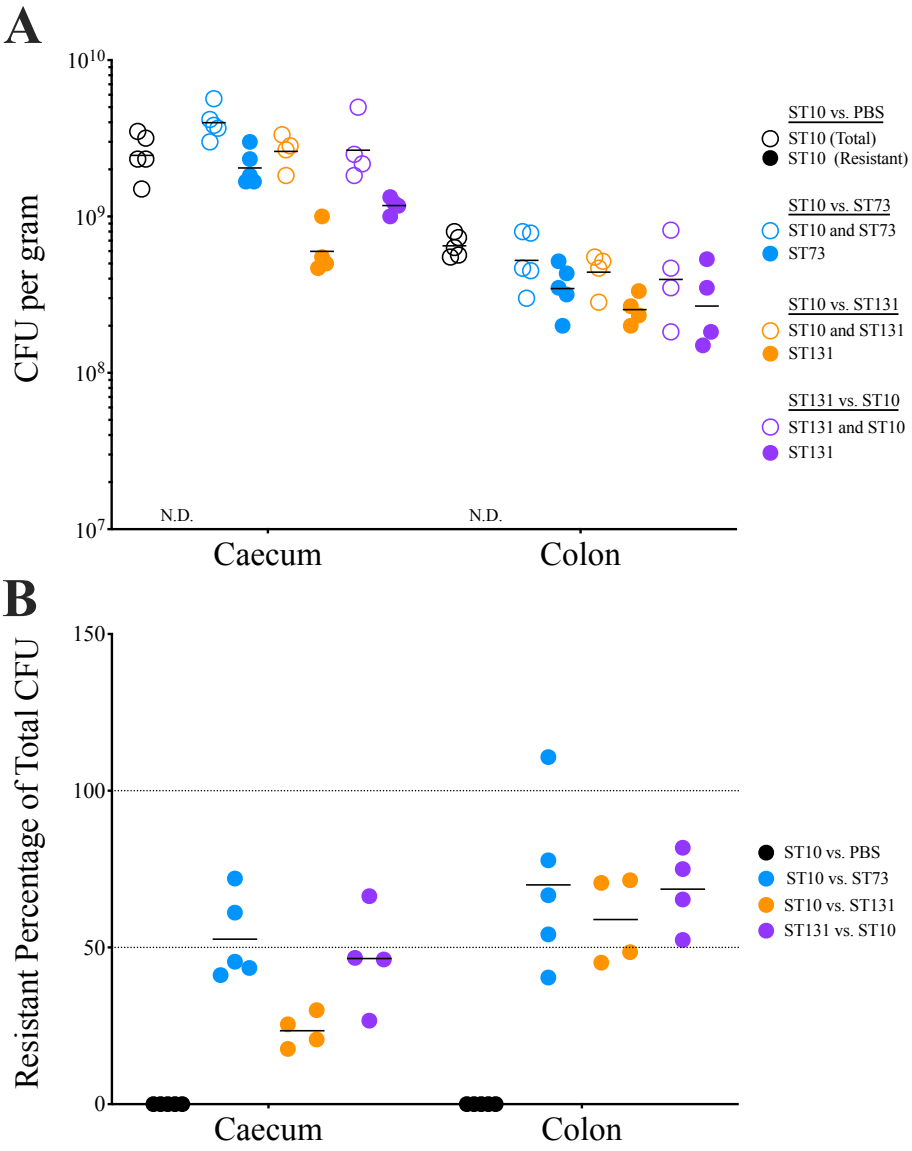

Fig. 5.16 CFU Values from Gut Contents of Challenged Mice. Points represent values from individual mice and are coloured by colonising strain with black representing ST10 vs. PBS, blue representing ST10 vs. ST73, orange representing ST10 vs. ST131 and purple representing ST131 vs. ST10. Open circles indicate total growth of both strains while closed circles indicate growth on selective agar. A) Raw CFU values per gram of gut contents. Lines indicate geometric mean. B) Percentage of total growth that is attributable to resistant strain, calculated by dividing growth on selective agar by total growth on non-selective agar. Lines indicate mean value.

Strain dissemination is seen frequently in ST131 colonised mice

Multiple tissues were harvested from the mice 3 weeks after challenge to determine whether the colonising strains had disseminated to other tissue sites (Figure 5.17, Table 5.1). Two mice that were commensally colonised and challenged with PBS showed dissemination to the MLN but bacteria were not detected in other tissues nor blood (Figure 5.17a). One mouse challenged with ST73 showed dissemination to the MLN, spleen and liver but no bacteria were detected in the blood (Figure 5.17b). In this instance the disseminating bacteria appeared to be predominantly the commensal with ST73 only accounting for 38.5% of growth in the spleen and 20% in the liver; while in the MLN 63.6% of the growth was from ST73. Two mice that were challenged with ST131 showed dissemination to the spleen and liver (Figure 5.17c). Of the mice that displayed dissemination to the spleen the majority was ST131 with 90.5% growth on selective agar compared to non-selective. Curiously one mouse showed growth on the selective agar but no growth on the non-selective agar. For the dissemination to the liver growth on selective agar was 62.1% of total growth for one mouse and 63.4% for the other, indicating that ST131 was the predominant disseminating strain but the commensal was also present. Lastly, the mice that were colonised with ST131 then challenged with ST10 showed the most dissemination with multiple mice having detectable bacteria in their MLN, spleen and liver (Figure 5.17d). Interestingly there were two mice that showed dissemination of the commensal only with no detectable ST131 growth from the spleen and liver of one mouse and the MLN of another mouse. Dissemination in the other mice was either completely ST131 or a mixture of both strains. This data suggests that mice that have been colonised with ST131 for longer show increased levels of bacteria in extra-intestinal sites.

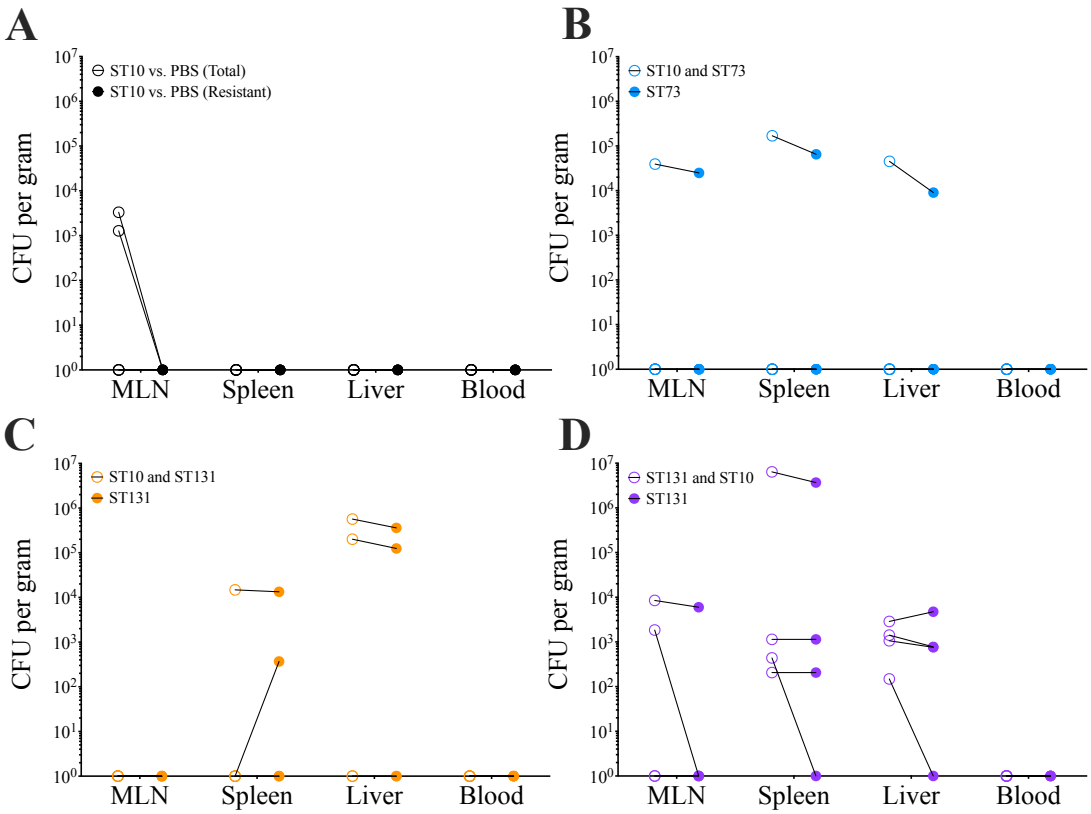

Fig. 5.17 CFU Values from Extra-Intestinal Tissues from Challenged Mice. Points represent CFU values from individual mice and are coloured by colonisation condition with black representing ST10 vs. PBS (A), blue representing ST10 vs. ST73 (B), orange representing ST10 vs. ST131 (C) and purple representing ST131 vs. ST10 (D). Open circles indicate total growth of both strains while closed circles indicate growth on selective agar. Lines connect CFU measurements from the same mouse.

Table 5.1 Percentage of growth on selective agar plates as compared to non-selective plates. Asterix indicates imputed value as there was no growth on non-selective plate.

Colonisation	Mouse	MLN	Spleen	Liver	Blood
ST10 challenged with ST73	1	-	-	-	-
	2	-	-	-	-
	3	-	-	-	-
	4	-	-	-	-
	5	63.6%	38.5%	20%	-
ST10 challenged with ST131	1	-	90.5%	62.2%	-
	2	-	-	-	-
	3	-	*100%**	63.4%	-
	4	-	-	-	-
ST131 challenged with ST10	1	0.0%	100.0%	164.2%	-
	2	-	57.5%	71.5%	-
	3	70.3%	100.0%	54.6%	-
	4	-	0.0%	0.0%	-

ST131 challenged mice show increased expression of inflammatory cytokines but no discernible inflammation by histology

As with previous experiments sections of the small intestine, caecum and colon were collected and RNA was isolated to determine cytokine mRNA expression. In the small intestine the expression of *S100A8* was significantly altered by the presence of bacteria with all colonisation conditions showing increased mRNA levels compared to germ-free control mice (Figure 5.18a). In the small intestine of mice colonised with ST131 then challenged with ST10 there was a significant increase in the expression of pro-inflammatory *CXCL1* and *IL-6* as well as anti-inflammatory *TGFβ* and *IL-10* when compared to germ-free mice. CXCL1 is produced from epithelial cells to recruit neutrophils while IL-6 and TFGβ influence the development of Th17 cells which can produce IL-10 to suppress inflammation [93]. It is unclear whether these mice are exhibiting the resolution of inflammation or localised regions of pro- and anti-inflammatory signalling. The small intestine of mice from other

challenge conditions show some significant differences in cytokine expression but this is limited to *IL-17*.

In the caecum several pro-inflammatory cytokines are up-regulated by the presence of bacteria (Figure 5.18b). *IL-17*, *CXCL1* and *S100A8* are all up-regulated under all colonisation conditions whilst *TNFα* and *IL-1β* are only upregulated when an ExPEC strain is present. The expression of IL-1β by resident immune cells is specifically increased by the presence of pathogenic bacteria rather than commensals [51]. The pro-inflammatory cytokines *TNFα*, *IL-1β*, *CXCL1* and *IL-6* are all significantly up-regulated when ST131 is invading compared to PBS challenged mice. This implies that when ST131 is invading an already colonised mice there is a stronger inflammatory response, which is not observed when ST10 nor ST73 is the invading strain. This is in keeping with the different colonisation dynamics displayed by the two experimental groups, specifically that ST131 displaces the commensal when it is invading but when the ST10 is invading the two strains exist in equivalent proportions.

The colon displays fewer alterations in cytokine response than the small intestine or caecum (Figure 5.18c). Both *TNFα* and *IL-17* are significantly increased when the mice are colonised by any strain. *IL-1β* and *CXCL1* are significantly up-regulated in mice challenged with ST10 compared to germ free. While *S100A8* is significantly up-regulated by mice challenged with ST131 compared to germ free.

Tissue sections from the small intestine and colon were examined for histological markers of inflammation. No differences were observed histologically, nor when scoring the tissue sections (Figure 5.19 & Figure 5.20). This is not unexpected as the colonised mice show small perturbations of cytokine expression in the small intestine and colon, suggesting that any inflammatory response is limited to the caecum.

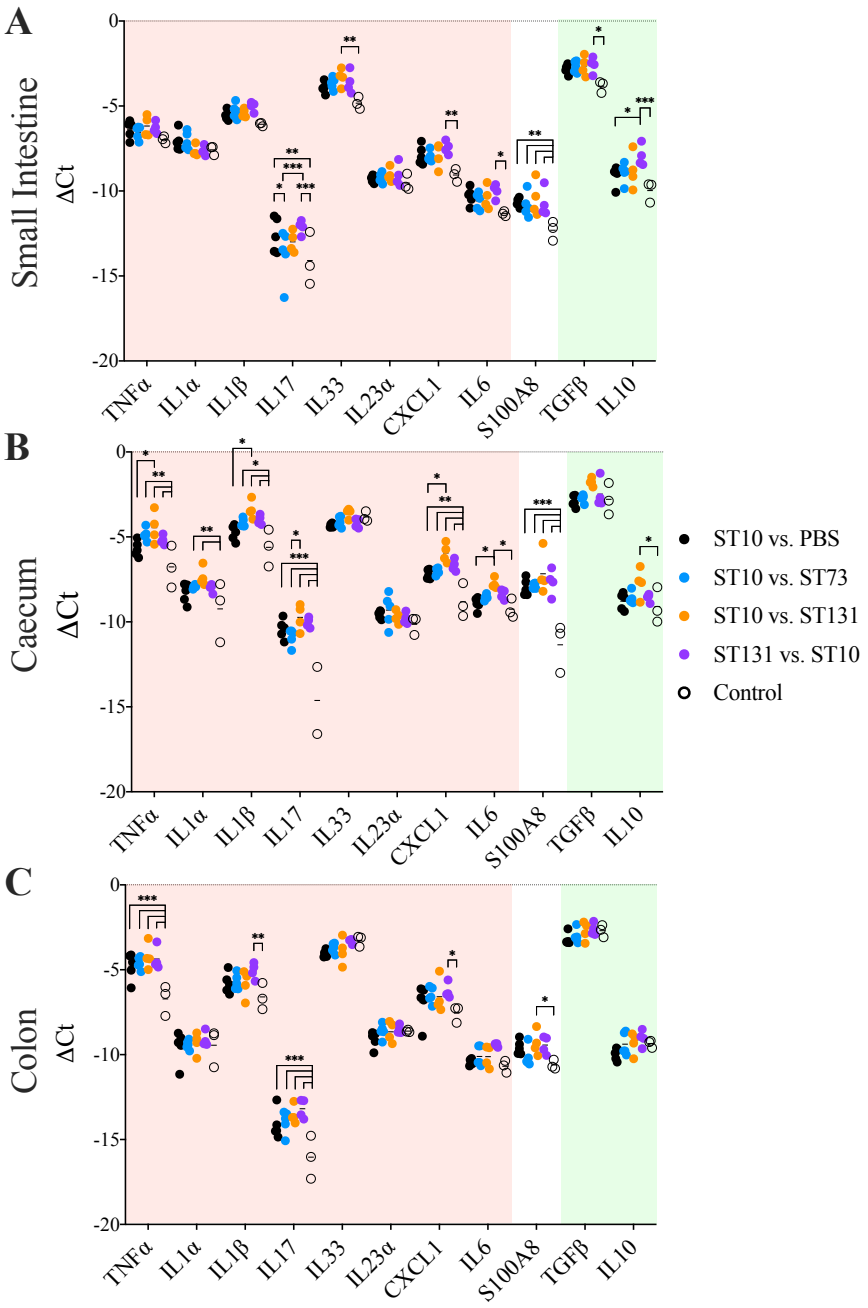

Fig. 5.18 Expression of cytokines assayed by probe based qPCR reveals significant increase in TNFα, IL-1β, CXCL1 and IL-6 in the caecum of ST131 challenged mice. Ct values are normalised to the housekeeping gene *Pol2ra* (ΔCt). Points indicate values for individual mice and are coloured according to colonisation (ST10 & ST73: blue, ST10 & ST131: orange, ST73 & ST131: pink, germ free: open circles). Lines indicate mean value. Cytokines are grouped into pro-inflammatory (red box) and anti-inflammatory (green box). Sections of the small intestine (A), caecum (B) and colon (C) were assayed. Significance determined by 2-way ANOVA with Tukey's multiple comparison, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. n=5.

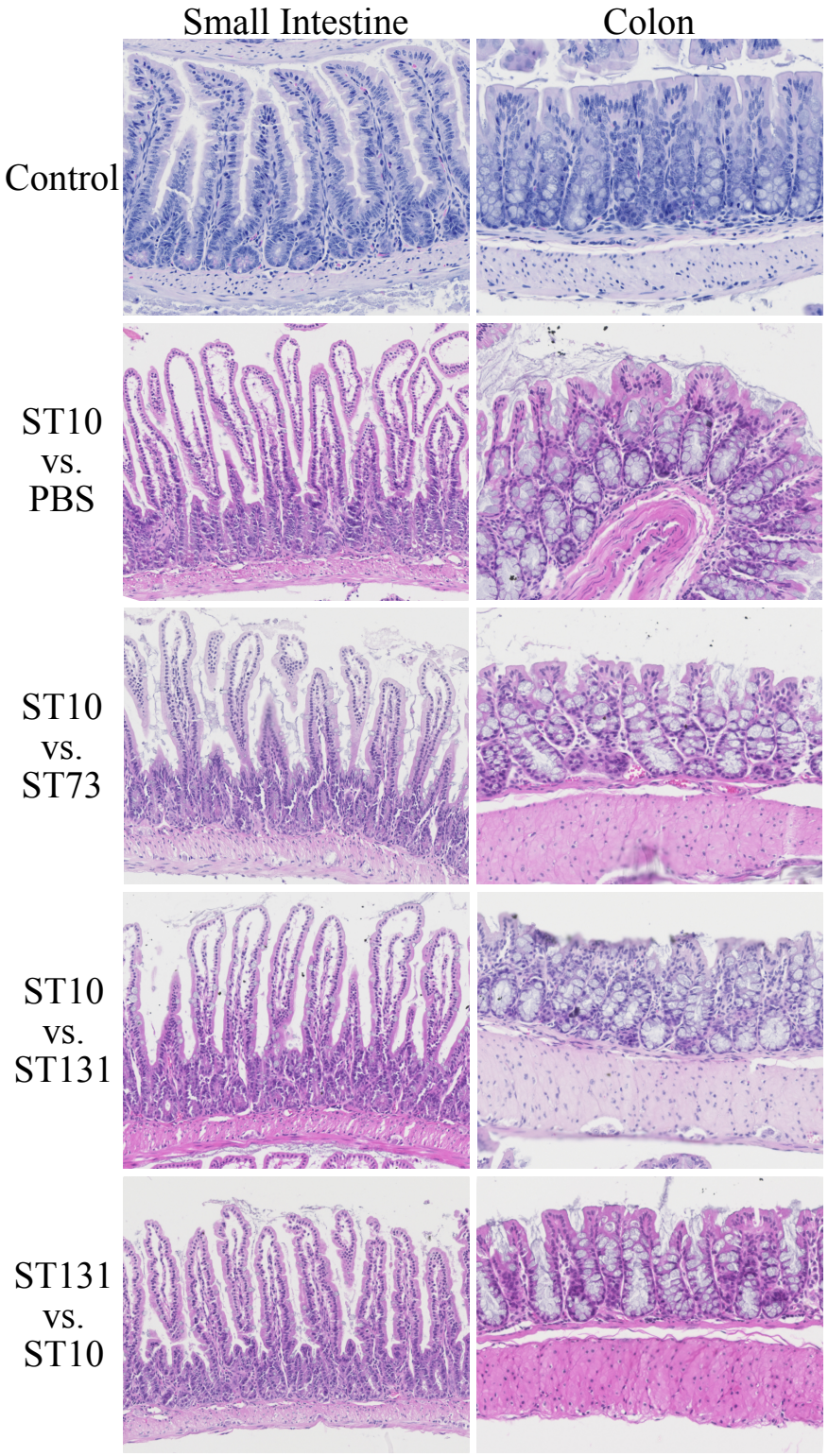

Fig. 5.19 Representative Tissue Sections of Small Intestine and Colon from Challenged Mice. Tissue sections have been H&E stained. Control group is non-colonised germ free mice.

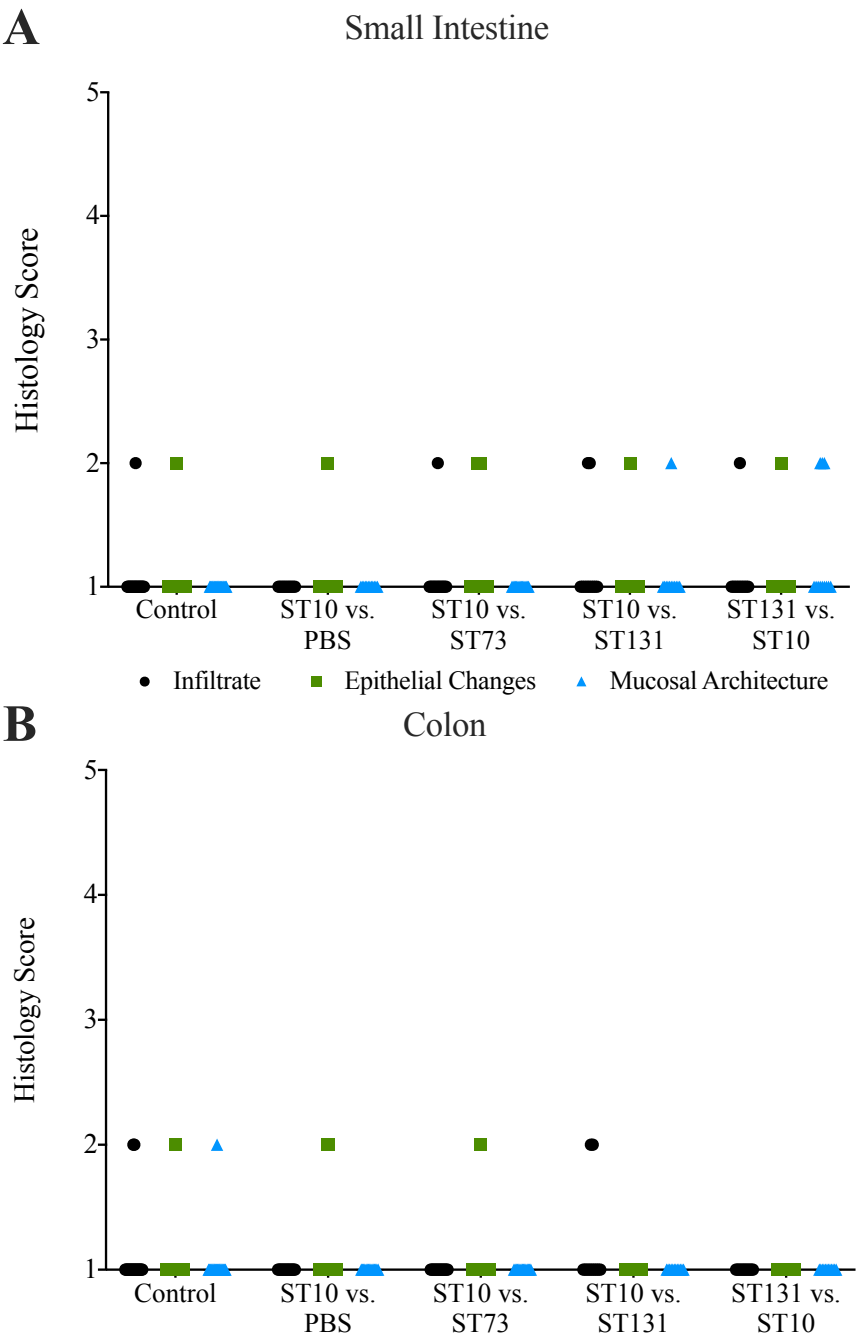

Fig. 5.20 Tissue Pathology Scores from Challenged Mice. Sections of small intestine (A) and colon (B) were blind scored using a scoring system for IBD mouse models. Scores range from 1 (absent / minor) to 5 (severe) for the categories: inflammatory infiltrate (black circles), changes to epithelium (green square) and alterations to overall mucosal architecture (blue triangles).

5.2.4 Transmission of ST131 Between Co-housed Mice

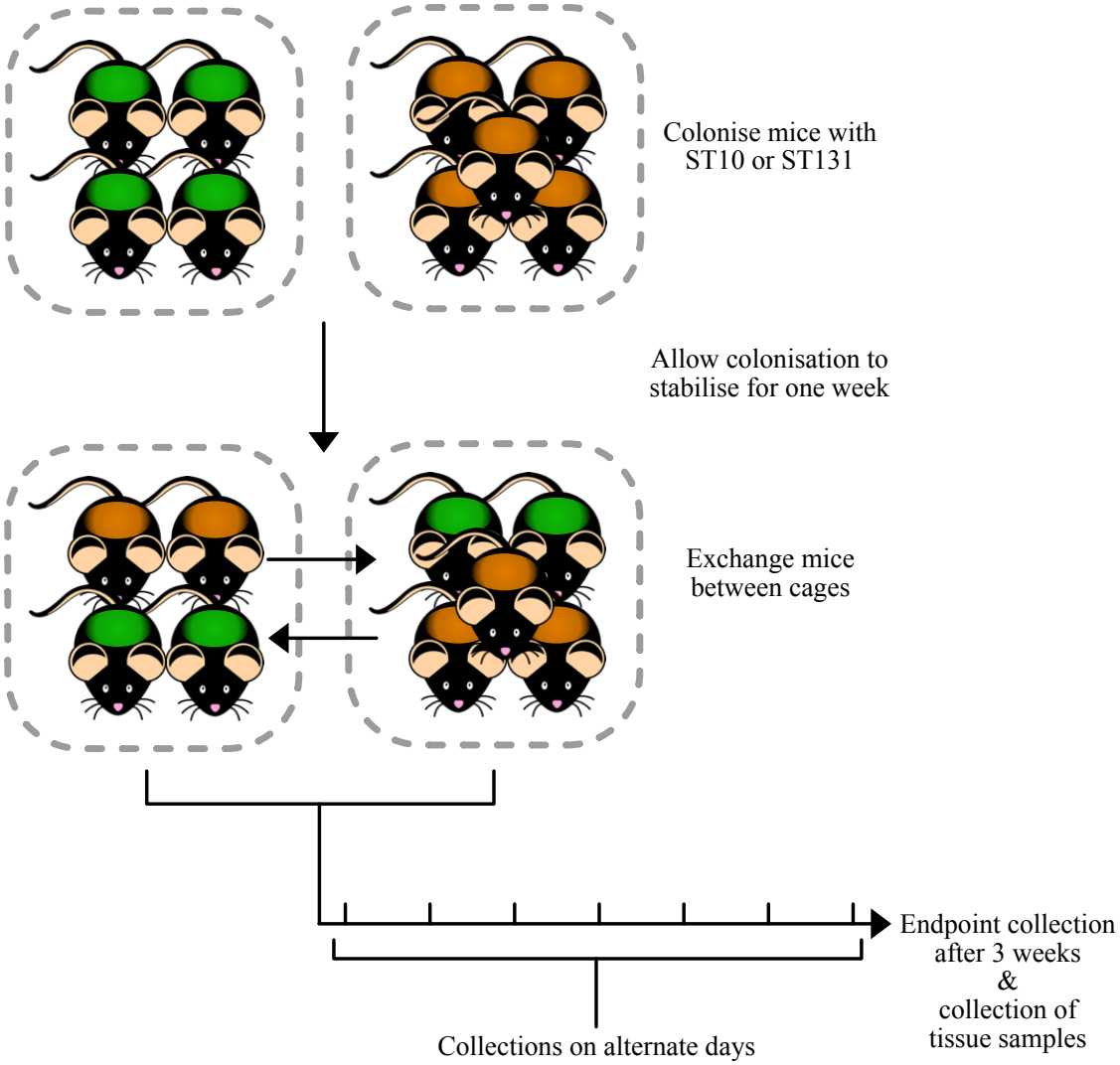

Fig. 5.21 Schematic of the colonisation of mice by either ST10 or ST131 for one week followed by exchange of mice between cages. Four mice were colonised with ST10 whilst 5 mice were colonised with ST131. Colonisation was allowed to stabilise before exchanging 2 mice between cages.

ST131 transmits between co-housed mice, becoming the dominant strain in all mice, in less than 48 hours

Gavaging mice is a highly artificial system of introducing the bacteria and not reflective of natural colonisation. To mimic a more natural colonisation scenario a co-housing methodology was used. For this mice were colonised with either ST10 or ST131 for one week before exchanging mice between cages (Figure 5.21). Colonisation was then monitored for 3 weeks post exchange by enumerating the CFU in faecal pellets. Under these conditions the mice that were gavaged with ST10 were colonised by ST131 within 2 days of co-housing (samples were only collected every 2 days so it is not possible to determine if colonisation occurred in a shorter time frame - Figure 5.22a). The colonisation by the invading ST131 followed a similar dynamic to that of the previous displacement experiments, specifically ST131 dominated over the resident commensal for the first 7 days before gradually reducing to equivalent proportions. Again when the ST131 is already a resident strain it does not out-compete the invading commensal, instead the two strains exist in equivalent proportions which are stable over the course of the experiment (Figure 5.22b). The invading ST10 does show a delay in its ability to colonise the host as it only appears at day 4 compared to ST131 which is apparent from day 2. These results are consistent with our previous experiment where a large amount of ST131 was artificially introduced. In this experiment ST131 was still an effective coloniser of ‘occupied’ mice even when introduced in small quantities.

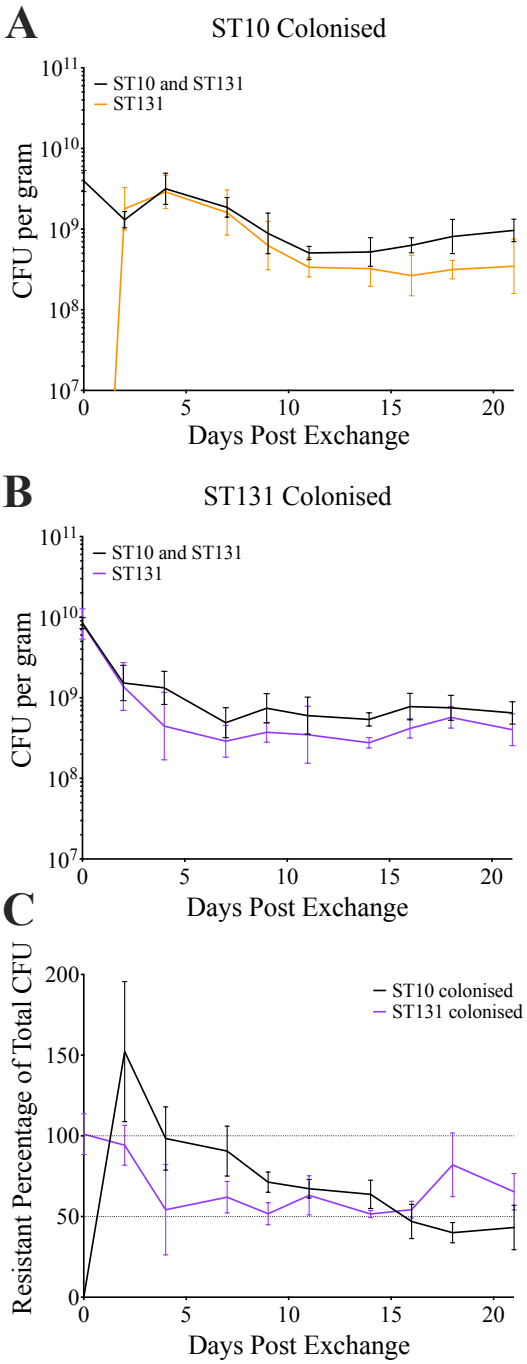

Fig. 5.22 Levels of CFU in the faecael pellets of co-housed mice. Lines indicate geometric mean, error bars indicate SD. Mice were co-housed at Day 0. A) Mice that were colonised with ST10. Black line indicates total growth of both strains (non-selective agar), orange line indicates growth of ST131 (selective agar). B) Mice that were colonsed with ST131. Black line indicates total growth of both strains (non-selective agar), purple line indicates growth of solely ST131 (selective agar). C) Summarisation of panels A and B. Percentage of growth that is attributable to resistant strain, calculated by dividing resistant CFU by total CFU. Black line indicates ST10 colonised mice, purple line indicates ST131 colonised mice. n=5.

Bacterial growth in the caecum is predominantly ST131 whilst there are varying levels of bacterial growth at extra-intestinal sites

After 3 weeks of co-housing the mice were sacrificed. Contents of the caecum and colon were collected and CFU values were enumerated. Levels of colonisation are still higher in the caecum than the colon, consistent with previous observations (Figure 5.23a). For the ST10 gavaged mice there was a total of 8.99×10^9 CFU/g of caecal contents of which 6.40×10^9 CFU/g was from ST131. For the ST131 gavaged mice there was a total growth of 9.90×10^9 CFU/g of caecal contents of which 7.65×10^9 CFU/g was ST131. From the contents of the colon of ST10 gavaged mice there was 3.78×10^9 CFU/g of which 2.08×10^9 CFU/g was ST131. For the ST131 gavaged mice there was 2.47×10^9 CFU/g of colon contents of which 1.54×10^9 CFU/g was ST131. From the ST10 gavaged mice the average growth attributable to the invading ST131 was 78.7% of the total growth from the caecum contents and 58.5% from the colon contents (Figure 5.23b). For the ST131 gavaged mice on average 81.5% of growth in the caecal contents was attributable to ST131 and from the colon contents it was 65.6%. In contrast to the previous experimental setup ST131 maintains dominance over the commensal in the caecum but the two strains exist in equivalent quantities in the colon. After 3 weeks of co-housing mice were sacrificed and multiple tissues were collected to measure dissemination (Figure 5.24, Table 5.2). The ST10 gavaged mice showed dissemination only to the spleen and MLN (Figure 5.24a). The disseminating bacteria were a mixture of ST10 and ST131. From the ST131 gavaged mice there were detectable bacteria in the MLN, spleen and liver (Figure 5.24b), which appeared to be either entirely commensal or ST131. As observed previously mice that are colonised with ST131 for longer more frequently have detectable bacteria in extra-intestinal sites.

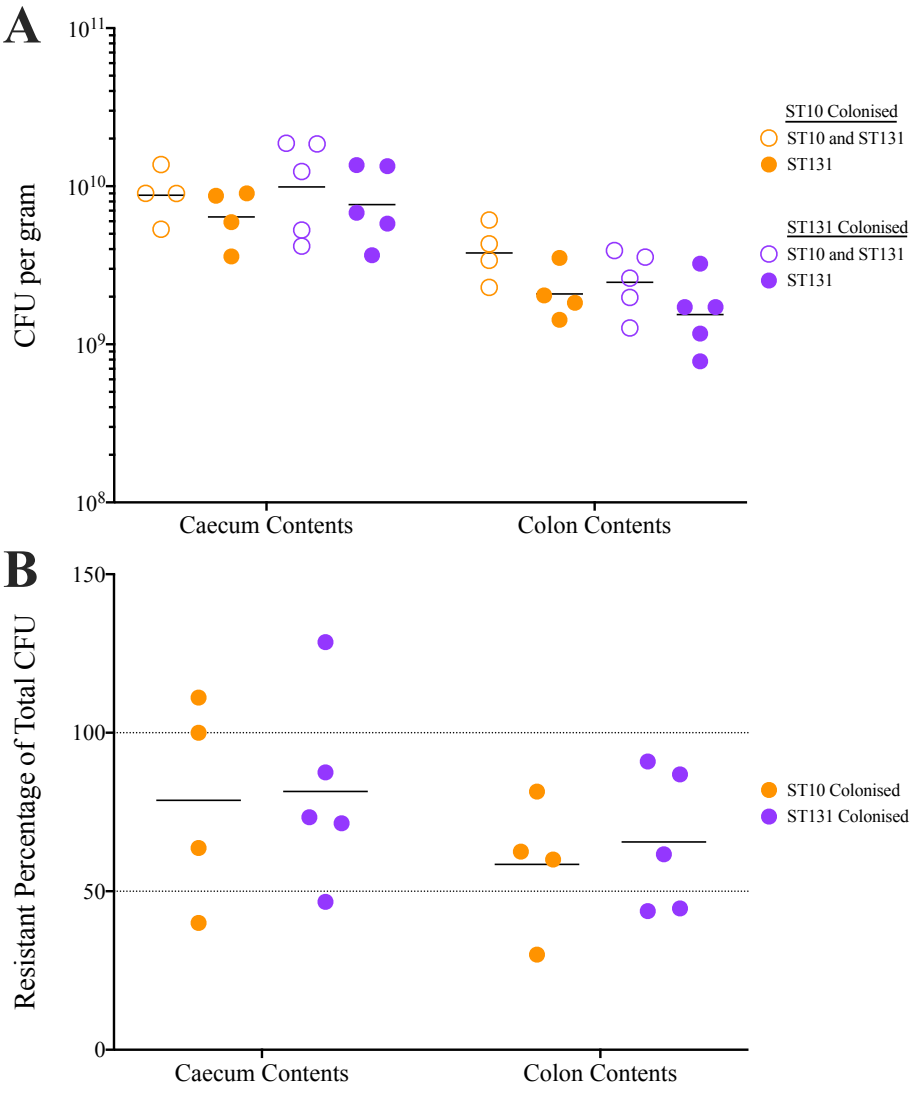

Fig. 5.23 CFU Values from Gut Contents of Co-housed Mice. Points represent values from individual mice and are coloured by colonising strain with orange representing ST10 colonised mice and purple representing ST131 colonised mice. Open circles indicate total growth of both strains while closed circles indicate growth on selective agar. A) Raw CFU values per gram of gut contents. Lines indicate geometric mean. B) Percentage of total growth that is attributable to resistant strain, calculated by dividing growth on selective agar by total growth on non-selective agar. Lines indicate mean value.

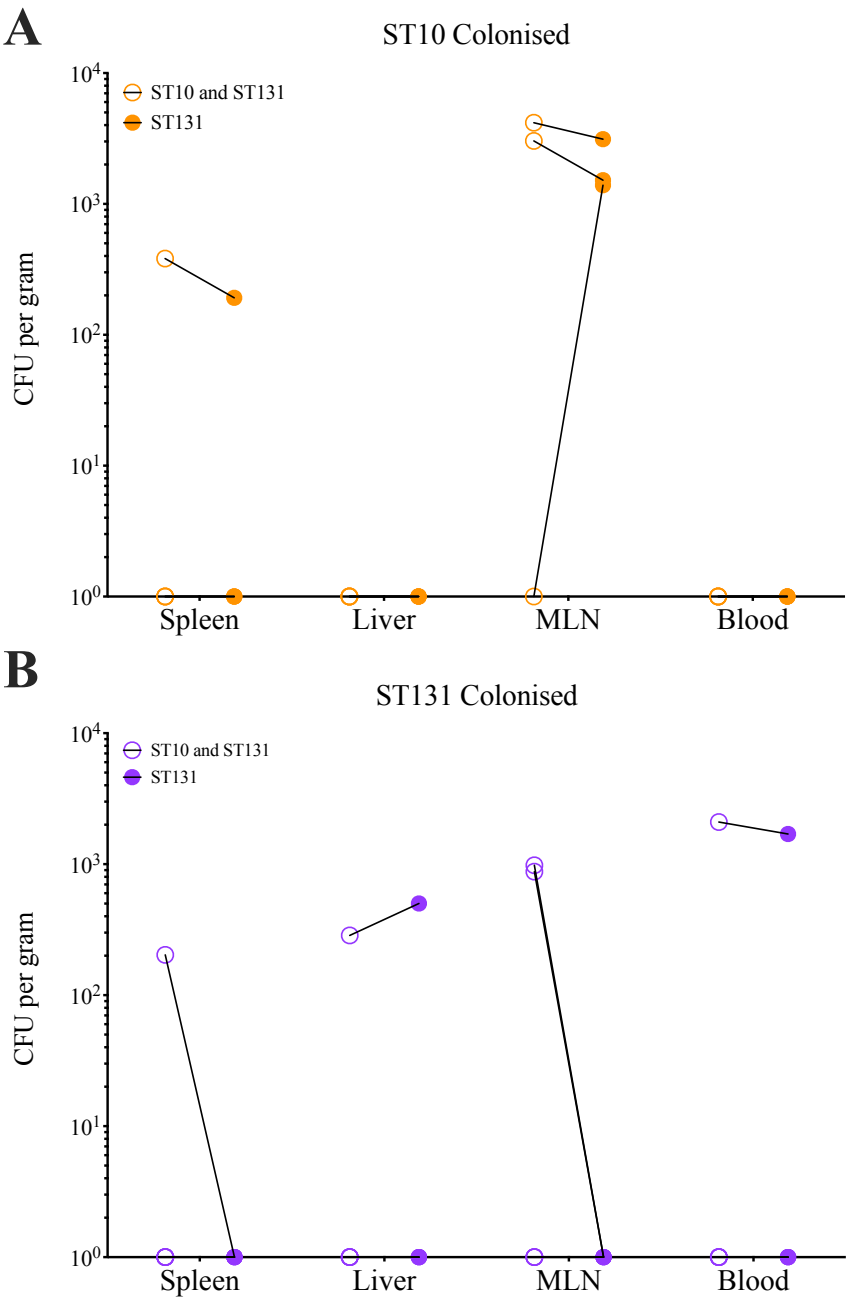

Fig. 5.24 CFU Values from Extra-Intestinal Tissues from Co-housed Mice. Points represent CFU values from individual mice and are coloured by colonisation condition with orange representing ST10 colonised mice (A) and purple representing ST131 colonised mice (B). Open circles indicate total growth of both strains while closed circles indicate growth on selective agar. Lines connect CFU measurements from the same mouse.

Table 5.2 Percentage of growth on selective agar plates as compared to non-selective plates. Asterix indicates imputed value as there was no growth on non-selective plate.

Colonisation	Mouse	MLN	Spleen	Liver	Blood
ST10 gavaged	1	-	-	50.2%	-
	2	*100%**	-	-	-
	3	-	-	-	-
	4	50.1%	-	75.1%	-
ST131 gavaged	1	-	-	-	81.0%
	2	-	-	0.0%	-
	3	0.0%	175.2%	0.0%	-
	4	-	-	-	-
	5	-	-	-	-

No observable difference in cytokine response between co-housed groups

The host cytokine response in the caecum was measured using qPCR (Figure 5.25). Previous experiments indicated that the inflammatory response was confined to the caecum, while the colon and small intestine displayed some significant alterations in cytokine expression they were generally produced by the presence of bacteria of any strain. Therefore we only examined the cytokine expression in the caecum of the co-housed mice. From the co-housed mice there was a significant difference in the expression of *S100A8* compared to germ-free mice. Curiously many of the cytokines that were significantly up-regulated in the presence of bacteria, regardless of ST, in the artificial challenge experiment are not altered in the co-housed mice. Specifically *CXCL1* is significantly increased in all groups in the previous challenge experiment when compared to germ-free, but levels of expression in the co-housed mice are comparable to germ-free. Again *IL-17* expression is increased in all mice in the challenge experiment but is unaffected in the co-housed mice.

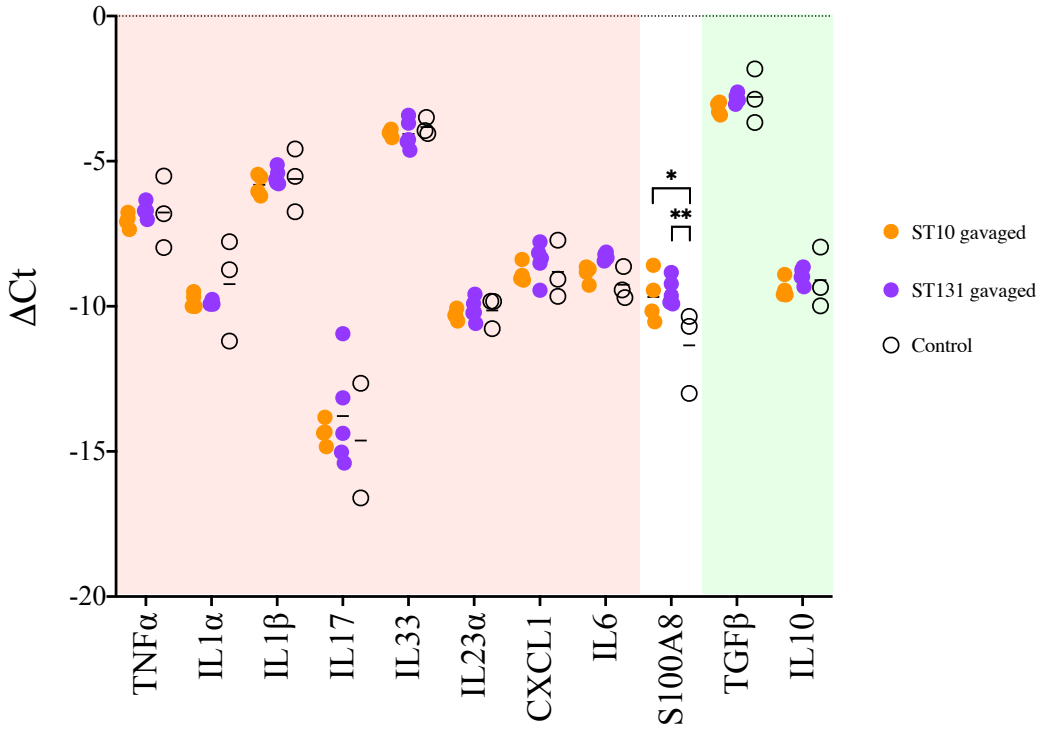

Fig. 5.25 Expression of multiple cytokines in the caecum of co-housed mice as assayed by probe based qPCR. Cycle threshold values are normalised to the housekeeping gene *pol2ra* (ΔCt). Points indicate values for individual mice and are coloured according to colonisation conditions with ST10 colonised mice in orange, ST131 colonised mice in purple and non-colonised germ free mice in open circles. Lines indicate mean value. Cytokines are grouped into those generally considered as pro-inflammatory (red box) and those considered anti-inflammatory (green box). Significance determined by 2-way ANOVA with Tukey's multiple comparison, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. n=5

5.3 Discussion

The data presented here indicate that ST131 is able to out-compete a commensal ST10 strain when introduced into a germ free mouse at the same time. This finding is mirrored by previous studies demonstrating that ST131 is able to out-compete commensal competitors when introduced into a mouse concurrently, however, this was achieved using a modified streptomycin treated mouse model [155]. We go on to demonstrate that an invading ST131 isolate is capable of displacing a resident commensal *E. coli* strain, both when the invading bacteria is introduced artificially via oral gavage and when mice are co-housed. In both instances this displacement occurs in the absence of antibiotic treatment, demonstrating that ST131 possesses some mechanism by which it can displace commensal *E. coli*. This is in keeping with recent observations that ST131 strain EC958 is able to colonise mice that have not been treated with streptomycin [133]. Moreover, ST131 has been observed to spread between humans with reported transmission within households [69]. In addtion MDR *E. coli* are the most frequent colonisers of healthy travellers able to colonise individuals in the absence of antibiotics [144]. Numerous studies of healthy individuals who travel to regions where antibiotic resistance is endemic have reported varying levels of colonisation of between 30% and 70% [163], however these studies rely on post travel sampling. A more recent study which samples healthy individuals throughout their period of travel observed a much higher colonisation rate of 95% [75]. *E. coli* was identified to be the most common colonising species but was only a transient coloniser with multiple individuals exhibiting a short period of colonisation. Our data fits with the more recent finding of high levels of colonisation, however we did not observe transient colonisation in this model. Long periods of MDR colonisation are infrequently reported by traveller studies, for example one study reported that 33.9% of colonised individuals were still colonised 1 month after travel [128].

From our data the resident commensal strain was not completely displaced and began to be detectable once more after a period of 7 to 10 days. Similar observations have been

made of human travellers, specifically individuals that travelled to India and were colonised by MDR *E. coli* had a recurrence of their original commensal *E. coli* upon their return [15]. It is interesting to note that the strains which persisted through the period of travel were predominantly ST131 while ST10 strains were replaced by other commensal strains. Our data mimics this observation in that a resident ST131 strain is not displaced by an invading ST10 strain. Moreover it implies that ST131 is inherently resistant to displacement either by a commensal or another MDR strain. We observe that the displacement of the commensal by ST131 lasts longer than the displacement by ST73, this mimics the observation made in humans where ST131 was observed to persist within a care-home for considerably longer than other ExPEC clones, such as ST73 [111].

Three weeks after mice were challenged with ST131 we detected increased mRNA levels of the pro-inflammatory cytokines *TNF*α, *IL1*β, *IL-6* and *CXCL1* in the caecum. Expression of TNFα by plasma cells has been implicated in pathogen clearance in mice [53], while IL-1β is only produced by mononuclear phagocytes upon exposure to pathogens and not commensals [51]. IL-1β has also been implicated as important for repair of the epithelial barrier following colitis [13]. CXCL1 is released from epithelial cells to recruit neutrophils to aid in pathogen clearance [139, 54, 117], lastly IL-6 has a diverse pro-inflammatory effects including effector cell differentiation [93, 117]. Together these cytokines suggest that the host is responding to an invading ST131 as a pathogen and that it is attempting to clear it from the intestine, potentially through plasma cells and neutrophils. Our observed increase in inflammatory cytokines is confined to the caecum, whilst the other assayed tissues displayed alterations in cytokine expression that were not unique to ST131. Previous work has demonstrated that ST131 colonisation burden, and that of *Salmonella*, is highest in the caecum compared to other compartments [133, 63], an observation we also make. Moreover our histology analysis of the small intestine and colon displayed no differences between colonisation conditions. This is consistent with previous work that observed no histopathological

differences in the colon of mice colonised with ST131 [133]. Examining the caecum for histological signs of inflammation would be of great interest. Overall our data indicate that there is an inflammatory response in the caecum when ST131 is an invading strain. To date there have been no investigations into the human immune response to colonisation by MDR bacteria however it has been reported that traveller's diarrhoea is the most frequently reported symptom by those who are colonised [76]. This is despite the majority of colonising strains being ExPEC pathotypes rather than InPEC, suggesting that there is a mild inflammatory response from the host following colonisation. Moreover, it has also been observed that travel, and colonisation by MDR bacteria, increases the abundance of Proteobacteria in the gut [12], this phylum is known to proliferate in inflamed guts [40].

Ultimately the data presented here demonstrates that ST131 is an effective coloniser of the mouse intestine even in the presence of competing strains of *E. coli*. ST131 is capable of rapidly displacing a resident commensal strain to become the dominant bacteria. We made this observation using an artificial oral gavage in which a large number of bacteria are introduced into the mouse but the same effect is also observed when mice are co-housed and a smaller number of bacteria transmit.

Chapter 6

Conclusions & Future Work

The increasing prevalence of antibiotic resistance is threatening health care systems. Previously treatable diseases are becoming increasingly challenging to treat as we run out of effective therapeutic strategies. Understanding the emergence and spread of antibiotic resistance is important if we wish to effectively combat it. There are still outstanding questions over how antibiotic resistance emerges and spreads. It has been documented, and shown by the data presented here, that antibiotic resistance is not equally prevalent across the *E. coli* population but rather concentrated in specific lineages. Moreover these lineages are able to spread between healthy individuals in the absence of antibiotic selection, implying that these lineages possess mechanisms for effectively colonising hosts. It is imperative we understand what factors are permitting these lineages to evolve such high levels of antibiotic resistance and spread globally. The work presented in this thesis attempts to understand what distinguishes MDR lineages from the rest of the *E. coli* population.

Genomic analysis of the *E. coli* population confirmed that antibiotic resistance is not equivalently distributed across all genomes. Instead specific lineages display an incredibly high average number of resistance genes per genome, moreover in excess of 90% of these populations carry multiple resistance genes. Further work could examine the prevalence of specific resistance genes in different populations. There is a wide variety of resistance

genes present in the population which confer varying levels of resistance to their host bacterium. For example the presence of a *tet* gene confers resistance to tetracycline while the presence of CTX-M confers resistance to multiple generations of β-lactam antibiotics. Currently, carbapenem antibiotics are considered the last line of defence against MDR pathogens remaining largely effective, however carbapenemase enzymes are beginning to emerge and are being spread by select clones of *E. coli* [172, 169, 50]. The analysis here could be further refined to focus on the prevalence of a manually curated database of the most clinically important resistance genes, such as those which confer resistance carbapenems.

Analysis of functionally annotated pangenomes revealed a significant association between carriage of antibiotic resistance genes and multiple metabolic categories. Namely lineages that display a high carriage of antibiotic resistance also show an increased abundance of energy production and conversion genes in their core genome. Further work could focus on comparisons between the core genomes of lineages to identify specific genes or pathways that are enriched in MDR lineages. Our pangenome analysis also identified that MDR lineages display increased variation in genes associated with carbohydrate metabolism and transport. It would be interesting to repeat this analysis on a genome-by-genome basis, specifically determining the functional composition of individual genomes as well as the number, or type, of resistance genes they harbour. Further investigation of the metabolic capability of MDR genomes would also be of interest, specifically using tools which are capable or predicting which metabolic reactions a bacterium can perform, based upon its genome, and correlating those against carriage of resistance. This could be further refined to focus on genomes which carry clinically important resistance genes, such as those conferring resistance to carbapenems.

Here we postulate that lineages are experiencing selection on metabolic categories to facilitate colonisation of new hosts which in turn may increase exposure to antibiotics and resistance genes. However, it is possible that bacteria are simply responding to the metabolic

burden imposed by the presence of antibiotic resistance genes and the plasmid upon which they are carried. Therefore it is important we determine the evolutionary sequence of events; does metabolic adaptation precede the acquisition of resistance genes. This is a challenging question to answer from a genomic perspective and may require the use of experimental evolution experiments.

We hypothesise that our observed metabolic signature associated with carriage of resistance genes is providing a colonisation advantage to MDR *E. coli* and that these lineages may be exploiting the host inflammatory response, as has been observed for *Salmonella*. To test this hypothesis, we used *in vitro* cell culture models to compare the competitive advantage of different strains of *E. coli*. These models failed to fully recapitulate the host response but did indicate that ST131 is able to out-compete commensal strains. To better model the host response we moved into a germ-free mouse model. From these experiments we observed that ST131 is able to displace a commensal strain both when artificially introduced into the mouse and in a co-housing scenario. Further work could determine the minimum infectious dose of ST131 in comparison to ST73, this could then be expanded to other MDR and non-MDR ExPEC lineages.

We observed that the order of colonisation affects the interaction between the commensal and ST131 in our mouse model. specifically ST131 is able to displace a resident commensal but does not prevent a commensal from colonising when ST131 is already established. These observations were made using CFU assays however there is a threshold of detection when using this technique, specifically the commensal strain may be present but at too low a concentration for us to detect. Using qPCR with strain specific primers would allow us to accurately quantify the amount of each strain present, even if they are several orders of magnitude different. Using this approach would also allow us to screen for our strain in the presence of a native microbiota. Including the native microbiome is important for extrapolating our results to the human population. In the first instance, monitoring strain colonisation

in an altered Schaedler flora (ASF) mouse model could be used as these mice are more representative of conventionally housed mice yet possess a microbiome with known diversity. ASF mice are colonised by only 8 bacterial species, specifically those which are found to be most abundant in conventionally housed mice. Importantly *E. coli* is absent in these mice, allowing us to introduce our strains of choice.

Our genomic population analysis indicates that carbohydrate metabolism is important to MDR lineages so it would be of great interest to determine the expression of metabolic genes in MDR strains as they are invading a host. To achieve this we could sample the gut contents at the peak of colonisation and measure gene expression by RNAseq. This could reveal new insights into how ST131 is able to displace commensals. In addition, our genomic data suggested that carbohydrate metabolism was important to MDR lineages, it would therefore be interesting to examine the effect of altering the carbohydrate levels in the diet has on the ability for MDR strains to colonise mice. Morevoer our genomic data identified the metabolic signature in multiple MDR lineages of *E. coli* yet we have only tested the ability of ST131 to colonise mice. Expanding these colonisation experiments, and gene expression assays, to other MDR lineages such as ST167 and ST648 would be of interest.

We hypothesised that MDR strains would induce a stronger inflammatory response from the host as they exploit this response to colonise the host. We observed that mice who had their commensal strain displaced by an invading ST131 exhibited significantly elevated expression of specific pro-inflammatory cytokines in the caecum 3 weeks after exposure to ST131. This was not observed in mice where a non-MDR ExPEC strain was introduced. However these observations were made 3 weeks after the introduction of ST131, this was past the peak of colonisation by ST131. Therefore it is important that samples of host gut tissue are collected at earlier timepoints, specifically within a 24 to 72 hour window post challenge. The host immune response could be determined using a qPCR approach, the same as for the data presented in Chapter 5. If the results from this are promising further

investigation could be conducted by RNAseq or isolation of specific immune cells from the samples, such as neutrophils, macrophages or T cells. Once isolated the activation state or phenotype of these immune cell populations can be determined using flow cytometry to assess how the host is responding.

Ultimately the work presented in this thesis demonstrates that antibiotic resistance carriage is not equivalent across the *E. coli* population but instead is concentrated in specific lineages. Moreover these lineages have an increased abundance of metabolic genes in their core genome alongside increased variation in carbohydrate metabolism. We go on to demonstrate that an ST131 strain is able to out-compete commensal strains in an *in vitro* cell culture model. The ST131 strain is also able to out-compete the commensal when introduced into a germ-free mouse. Moreover the ST131 is able to displace a commensal that has already established itself in the host, this is observed when the ST131 is artificially introduced as well as when the mice are co-housed and transmission occurs naturally. This work contributes to a greater understanding of the evolution of antibiotic resistance and its transmission.

References

[1] Alexa, A. and Rahnenfuhrer, J. (2016). topGO: enrichment analysis for Gene Ontology. R Packag. version 2.26.0. *R Package version 2.26.0*.

[2] Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z., and McNally, A. (2016). Increase in bacteraemia cases in the East Midlands region of the UK due to MDR Escherichia coli ST73: High levels of genomic and plasmid diversity in causative isolates. *Journal of Antimicrobial Chemotherapy*, 71(2):339–343.

[3] Alhashash, F., Weston, V., Diggle, M., and McNally, A. (2013). Multidrug-resistant Escherichia coli bacteremia. *Emerging Infectious Diseases*, 19(10):1699–1701.

[4] Alqasim, A., Emes, R., Clark, G., Newcombe, J., La Ragione, R., and McNally, A. (2014a). Phenotypic microarrays suggest Escherichia coli ST131 is not a metabolically distinct lineage of extra-intestinal pathogenic E. coli. *PLoS ONE*, 9(2).

[5] Alqasim, A., Jaffal, A. A., Almutairi, N., and Alyousef, A. A. (2021). Comparative phenotypic characterization identifies few differences in the metabolic capacity between Escherichia coli ST131 subclones. *Saudi Journal of Biological Sciences*, 28(1):762–769.

[6] Alqasim, A., Scheutz, F., Zong, Z., and McNally, A. (2014b). Comparative genome analysis identifies few traits unique to the Escherichia coli ST131 H30Rx clade and extensive mosaicism at the capsule locus. *BMC Genomics*, 15(1):1–8.

[7] Arcilla, M. S., van Hattem, J. M., Haverkate, M. R., Bootsma, M. C., van Genderen, P. J., Goorhuis, A., Grobusch, M. P., Lashof, A. M., Molhoek, N., Schultsz, C., Stobberingh, E. E., Verbrugh, H. A., de Jong, M. D., Melles, D. C., and Penders, J. (2017). Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. *The Lancet Infectious Diseases*, 17(1):78–85.

[8] Bandyopadhaya, A., Sarkar, M., and Chaudhuri, K. (2007). Transcriptional upregulation of inflammatory cytokines in human intestinal epithelial cells following Vibrio cholerae infection. *FEBS Journal*, 274(17):4631–4642.

[9] Banerjee, R. and Johnson, J. R. (2014). A new clone sweeps clean: The enigmatic emergence of Escherichia coli sequence type 131. *Antimicrobial Agents and Chemotherapy*, 58(9):4997–5004.

[10] Barrios, H., Garza-Ramos, U., Mejia-Miranda, I., Reyna-Flores, F., Sánchez-Pérez, A., Mosqueda-García, D., and Silva-Sanchez, J. (2017). ESBL-producing Escherichia coli

and Klebsiella pneumoniae: The most prevalent clinical isolates obtained between 2005 and 2012 in Mexico. *Journal of Global Antimicrobial Resistance*, 10:243–246.

[11] Ben Zakour, N. L., Alsheikh-Hussain, A. S., Ashcroft, M. M., Khanh Nhu, N. T., Roberts, L. W., Stanton-Cook, M., Schembri, M. A., and Beatsona, S. A. (2016). Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of Escherichia coli ST131. *mBio*, 7(2):1–11.

[12] Bengtsson-Palme, J., Angelin, M., Huss, M., Kjellqvist, S., Kristiansson, E., Palmgren, H., Joakim Larsson, D. G., and Johansson, A. (2015). The human gut microbiome as a transporter of antibiotic resistance genes between continents. *Antimicrobial Agents and Chemotherapy*, 59(10):6551–6560.

[13] Bersudsky, M., Luski, L., Fishman, D., White, R. M., Ziv-Sokolovskaya, N., Dotan, S., Rider, P., Kaplanov, I., Aychek, T., Dinarello, C. A., Apte, R. N., and Voronov, E. (2014). Non-redundant properties of IL-1α and IL-1β during acute colon inflammation in mice. *Gut*, 63(4):598–609.

[14] Bessa, J., Meyer, C. A., de Vera Mudry, M. C., Schlicht, S., Smith, S. H., Iglesias, A., and Cote-Sierra, J. (2014). Altered subcellular localization of IL-33 leads to non-resolving lethal inflammation. *Journal of Autoimmunity*, 55(1):33–41.

[15] Bevan, E. R., McNally, A., Thomas, C. M., Piddock, L. J. V., and Hawkey, P. M. (2018). Acquisition and Loss of CTX-M-Producing and Non-Producing Escherichia coli in the Fecal Microbiome of Travelers to South Asia. *mBio*, 9(6):1–14.

[16] Bonnet, R. (2004). Growing Group of Extended-Spectrum β-Lactamases: the CTX-M Enzymes. *Antimicrobial Agents and Chemotherapy*, 48(1):1–14.

[17] 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., Nykäsenoja, 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(12):3491–3500.

[18] Bottery, M. J., Wood, A. J., and Brockhurst, M. A. (2017). Adaptive modulation of antibiotic resistance through intragenomic coevolution. *Nature Ecology and Evolution*, 1(9):1364–1369.

[19] Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C. H., Xie, D., Zhang, C., Stadler, T., and Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4):1–28.

[20] Brolund, A., Edquist, P. J., Mäkitalo, B., Olsson-Liljequist, B., Söderblom, T., Wisell, K. T., and Giske, C. G. (2014). Epidemiology of extended-spectrum β-lactamase-producing Escherichia coli in Sweden 2007-2011. *Clinical Microbiology and Infection*, 20(6).

[21] Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1):59–60.

[22] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):421.

[23] Cantón, R. and Coque, T. M. (2006). The CTX-M β-lactamase pandemic. *Current Opinion in Microbiology*, 9(5):466–475.

[24] Cartelle, M., Del Mar Tomas, M., Molina, F., Moure, R., Villanueva, R., and Bou, G. (2004). High-level resistance to ceftazidime conferred by a novel enzyme, CTX-M-32, derived from CTX-M-1 through a single Asp240-Gly substitution. *Antimicrobial Agents and Chemotherapy*, 48(6):2308–2313.

[25] Cassini, A., Högberg, L. D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G. S., Colomb-Cotinat, M., Kretzschmar, M. E., Devleesschauwer, B., Cecchini, M., Ouakrim, D. A., Oliveira, T. C., Struelens, M. J., Suetens, C., Monnet, D. L., Strauss, R., Mertens, K., Struyf, T., Catry, B., Latour, K., Ivanov, I. N., Dobreva, E. G., Tambic Andraševic, A., Soprek, S., Budimir, A., Paphitou, N., Žemlicková, H., Schytte Olsen, S., Wolff Sönksen, U., Märtin, P., Ivanova, M., Lyytikäinen, O., Jalava, J., Coignard, B., Eckmanns, T., Abu Sin, M., Haller, S., Daikos, G. L., Gikas, A., Tsiodras, S., Kontopidou, F., Tóth, Á., Hajdu, Á., Guólaugsson, Ó., Kristinsson, K. G., Murchan, S., Burns, K., Pezzotti, P., Gagliotti, C., Dumpis, U., Liuimiene, A., Perrin, M., Borg, M. A., de Greeff, S. C., Monen, J. C., Koek, M. B., Elstrøm, P., Zabicka, D., Deptula, A., Hryniewicz, W., Caniça, M., Nogueira, P. J., Fernandes, P. A., Manageiro, V., Popescu, G. A., Serban, R. I., Schréterová, E., Litvová, S., Štefkovicová, M., Kolman, J., Klavs, I., Korošec, A., Aracil, B., Asensio, A., Pérez-Vázquez, M., Billström, H., Larsson, S., Reilly, J. S., Johnson, A., and Hopkins, S. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*, 19(1):56–66.

[26] Chang, D. E., Smalley, D. J., Tucker, D. L., Leatham, M. P., Norris, W. E., Stevenson, S. J., Anderson, A. B., Grissom, J. E., Laux, D. C., Cohen, P. S., and Conway, T. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19):7427–7432.

[27] Chen, C., Blumentritt, C. A., Curtis, M. M., Sperandio, V., Torres, A. G., and Dudley, E. G. (2013). Restrictive Streptomycin Resistance Mutations Decrease the Formation of Attaching and Effacing Lesions in Escherichia coli O157:H7 Strains. *Antimicrobial Agents and Chemotherapy*, 57(9):4260–4266.

[28] Clermont, O., Gordon, D., and Denamur, E. (2015). Guide to the various phylogenetic classification schemes for Escherichia coli and the correspondence among schemes. *Microbiology (Reading, England)*, 161(2015):980–988.

[29] Coburn, B., Li, Y., Owen, D., Vallance, B. A., and Finlay, B. B. (2005). Salmonella enterica serovar Typhimurium pathogenicity island 2 is necessary for complete virulence in a mouse model of infectious enterocolitis. *Infection and Immunity*, 73(6):3219–3227.

[30] Coque, T. M., Novais, Â., Carattoli, A., Poirel, L., Pitout, J., Peixe, L., Baquero, F., Cantón, R., and Nordmann, P. (2008). Dissemination of Clonally Related Escherichia coli Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15. *Emerging Infectious Diseases*, 14(2):195–200.

[31] Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J., and Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research*, 43(3):e15.

[32] Croxall, G., Hale, J., Weston, V., Manning, G., Cheetham, P., Achtman, M., and Mcnally, A. (2011). Molecular epidemiology of extraintestinal pathogenic Escherichia coli isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. *Journal of Antimicrobial Chemotherapy*, 66(11):2501–2508.

[33] Croxen, M. A. and Finlay, B. B. (2010). Molecular mechanisms of Escherichia coli pathogenicity. *Nature Reviews Microbiology*, 8(1):26–38.

[34] Cullen, I. M., Manecksha, R. P., McCullagh, E., Ahmad, S., O'Kelly, F., Flynn, R. J., McDermott, T., Murphy, P., Grainger, R., Fennell, J. P., and Thornhill, J. A. (2012). The changing pattern of antimicrobial resistance within 42 033 Escherichia coli isolates from nosocomial, community and urology patient-specific urinary tract infections, Dublin, 1999-2009. *BJU International*, 109(8):1198–1206.

[35] Darfeuille-Michaud, A., Boudeau, J., Bulois, P., Neut, C., Glasser, A. L., Barnich, N., Bringer, M. A., Swidsinski, A., Beaugerie, L., and Colombel, J. F. (2004). High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology*, 127(2):412–421.

[36] de Kraker, M. E., Jarlier, V., Monen, J. C., Heuer, O. E., van de Sande, N., and Grundmann, H. (2013). The changing epidemiology of bacteraemias in Europe: Trends from the European antimicrobial resistance surveillance system. *Clinical Microbiology and Infection*, 19(9):860–868.

[37] Denamur, E., Clermont, O., Bonacorsi, S., and Gordon, D. (2021). The population genetics of pathogenic Escherichia coli. *Nature Reviews Microbiology*, 19(1):37–54.

[38] Desvaux, M., Dalmasso, G., Beyrouthy, R., Barnich, N., Delmas, J., and Bonnet, R. (2020). Pathogenicity Factors of Genomic Islands in Intestinal and Extraintestinal Escherichia coli. *Frontiers in Microbiology*, 11(September).

[39] Diene, S. M. and Rolain, J. M. (2014). Carbapenemase genes and genetic platforms in Gram-negative bacilli: Enterobacteriaceae, Pseudomonas and Acinetobacter species. *Clinical Microbiology and Infection*, 20(9):831–838.

[40] Dinh, D. M., Volpe, G. E., Duffalo, C., Bhalchandra, S., Tai, A. K., Kane, A. V., Wanke, C. A., and Ward, H. D. (2015). Intestinal Microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *Journal of Infectious Diseases*, 211(1):19–27.

[41] Dogan, B., Suzuki, H., Herlekar, D., Sartor, B. R. B., Campbell, B. J., Roberts, C. L., Stewart, K., Scherl, E. J., Araz, Y., Bitar, P. P., Lefébure, T., Chandler, B., Schukken, Y. H., Stanhope, M. J., and Simpson, K. W. (2014). Inflammation-associated adherent-invasive escherichia coli are enriched in pathways for use of propanediol and iron and M-cell translocation. *Inflammatory Bowel Diseases*, 20(11):1919–1932.

[42] Dunn, S., Carrilero, L., Brockhurst, M., and McNally, A. (2021). Limited and Strain-Specific Transcriptional and Growth Responses to Acquisition of a Multidrug Resistance Plasmid in Genetically Diverse Escherichia coli Lineages. *mSystems*, 6(2).

[43] Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Microbiology: Diversity of the human intestinal microbial flora. *Science*, 308(5728):1635–1638.

[44] Engelsöy, U., Rangel, I., and Demirel, I. (2019). Impact of Proinflammatory Cytokines on the Virulence of Uropathogenic Escherichia coli. *Frontiers in Microbiology*, 10(MAY):1–12.

[45] Erben, U., Loddenkemper, C., Doerfel, K., Spieckermann, S., Haller, D., Heimesaat, M. M., Zeitz, M., Siegmund, B., and Kühl, A. A. (2014). A guide to histomorphological evaluation of intestinal inflammation in mouse models. *International journal of clinical and experimental pathology*, 7(8):4557–76.

[46] Erjavec, M. S. and Žgur-Bertok, D. (2015). Virulence potential for extraintestinal infections among commensal Escherichia coli isolated from healthy humans-the Trojan horse within our gut. *FEMS Microbiology Letters*, 362(5).

[47] European Centre for Disease Prevention and Control (2020). Antimicrobial Resistance in the EU/EEA (EARS-Net). Annual Epidemiological Report for 2019. *Surveillance report*.

[48] Faber, F., Thiennimitr, P., Spiga, L., Byndloss, M. X., Litvak, Y., Lawhon, S., Andrews-Polymenis, H. L., Winter, S. E., and Bäumler, A. J. (2017). Respiration of Microbiota-Derived 1,2-propanediol Drives Salmonella Expansion during Colitis. *PLoS Pathogens*, 13(1):1–19.

[49] 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(3):1143–1152.

[50] Feng, Y., Liu, L., Lin, J., Ma, K., Long, H., Wei, L., Xie, Y., McNally, A., and Zong, Z. (2019). Key evolutionary events in the emergence of a globally disseminated, carbapenem resistant clone in the Escherichia coli ST410 lineage. *Communications Biology*, 2(1):1–13.

[51] Franchi, L., Kamada, N., Nakamura, Y., Burberry, A., Kuffa, P., Suzuki, S., Shaw, M. H., Kim, Y. G., and Núñez, G. (2012). NLRC4-driven production of IL-1β discriminates between pathogenic and commensal bacteria and promotes host intestinal defense. *Nature Immunology*, 13(5):449–456.

[52] Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., an der Heiden, M., Bernard, H., Fruth, A., Prager, R., Spode, A., Wadl, M., Zoufaly, A., Jordan, S., Kemper, M. J., Follin, P., Müller, L., King, L. A., Rosner, B., Buchholz, U., Stark, K., and Krause, G. (2011). Epidemic Profile of Shiga-Toxin–Producing Escherichia coli O104:H4 Outbreak in Germany. *New England Journal of Medicine*, 365(19):1771–1780.

[53] Fritz, J. H., Rojas, O. L., Simard, N., McCarthy, D. D., Hapfelmeier, S., Rubino, S., Robertson, S. J., Larijani, M., Gosselin, J., Ivanov, I. I., Martin, A., Casellas, R., Philpott, D. J., Girardin, S. E., McCoy, K. D., MacPherson, A. J., Paige, C. J., and Gommerman, J. L. (2012). Acquisition of a multifunctional IgA + plasma cell phenotype in the gut. *Nature*, 481(7380):199–205.

[54] Gaffen, S. L., Jain, R., Garg, A. V., and Cua, D. J. (2014). The IL-23-IL-17 immune axis: From mechanisms to therapeutic testing. *Nature Reviews Immunology*, 14(9):585–600.

[55] Gibreel, T. M., Dodgson, A. R., Cheesbrough, J., Bolton, F. J., Fox, A. J., and Upton, M. (2012a). High metabolic potential may contribute to the success of ST131 uropathogenic Escherichia coli. *Journal of Clinical Microbiology*, 50(10):3202–3207.

[56] Gibreel, T. M., Dodgson, A. R., Cheesbrough, J., Fox, A. J., Bolton, F. J., and Upton, M. (2012b). Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. *Journal of Antimicrobial Chemotherapy*, 67(2):346–356.

[57] Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., and Harris, S. R. (2018). Phandango: An interactive viewer for bacterial population genomics. *Bioinformatics*, 34(2):292–293.

[58] Hall, B. G. and Barlow, M. (2004). Evolution of the serine β-lactamases: Past, present and future. *Drug Resistance Updates*, 7(2):111–123.

[59] Hammerum, A. M., Porsbo, L. J., Hansen, F., Roer, L., Kaya, H., Henius, A., Møller, K. L., Justesen, U. S., Søes, L., Røder, B. L., Thomsen, P. K., Wang, M., Søndergaard, T. S., Holzknecht, B. J., Østergaard, C., Kjerulf, A., Kristensen, B., and Hasman, H. (2020). Surveillance of OXA-244-producing Escherichia coli and epidemiologic investigation of cases, Denmark, January 2016 to August 2019. *Eurosurveillance*, 25(18):1–9.

[60] Harris, S. R. (2018). SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. *bioRxiv*, page 453142.

[61] Harrison, E., Guymer, D., Spiers, A. J., Paterson, S., and Brockhurst, M. A. (2015). Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Current Biology*, 25(15):2034–2039.

[62] Hazen, T. H., Leonard, S. R., Lampel, K. A., Lacher, D. W., Maurelli, A. T., and Raskoa, D. A. (2016). Investigating the relatedness of enteroinvasive Escherichia coli to Other E. coli and Shigella isolates by using comparative genomics. *Infection and Immunity*, 84(8):2362–2371.

[63] Hohmann, A. W., Schmidt, G., and Rowley, D. (1978). Intestinal Colonization and Virulence of Salmonella in Mice. *Infection and Immunity*, 22(3):763–770.

[64] Horesh, G., Blackwell, G. A., Tonkin-Hill, G., Corander, J., Heinz, E., and Thomson, N. R. (2021). A comprehensive and high-quality collection of escherichia coli genomes and their genes. *Microbial Genomics*, 7(2):1–15.

[65] Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8):2115–2122.

[66] Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., Von Mering, C., and Bork, P. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1):D286–D293.

[67] Irrgang, A., Falgenhauer, L., Fischer, J., Ghosh, H., Guiral, E., Guerra, B., Schmoger, S., Imirzalioglu, C., Chakraborty, T., Hammerl, J. A., and Käsbohrer, A. (2017). CTX-M-15-producing E. coli isolates from food products in Germany are mainly associated with an IncF-type plasmid and belong to two predominant clonal E. coli lineages. *Frontiers in Microbiology*, 8(NOV).

[68] Jiang, Z.-D., Greenberg, D., Nataro, J. P., Steffen, R., and DuPont, H. L. (2002). Rate of Occurrence and Pathogenic Effect of Enteroaggregative Escherichia coli Virulence Factors in International Travelers. *Journal of Clinical Microbiology*, 40(11):4185–4190.

[69] 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(12):4078–4082.

[70] Johnson, J. R., Johnston, B., Clabots, C., Kuskowski, M. A., and Castanheira, M. (2010a). Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant E. coli infections in the United States. *Clinical Infectious Diseases*, 51(3):286–294.

[71] Johnson, J. R., Johnston, B., Clabots, C., Kuskowski, M. A., Pendyala, S., DeBroy, C., Nowicki, B., and Rice, J. (2010b). Escherichia coli sequence type ST131 as an emerging fluoroquinolone- resistant uropathogen among renal transplant recipients. *Antimicrobial Agents and Chemotherapy*, 54(1):546–550.

[72] Johnson, J. R., Porter, S. B., Zhanel, G., Kuskowski, M. A., and Denamur, E. (2012). Virulence of Escherichia coli clinical isolates in a murine sepsis model in relation to sequence type ST131 status, fluoroquinolone resistance, and virulence genotype. *Infection and Immunity*, 80(4):1554–1562.

[73] Johnson, J. R., Tchesnokova, V., Johnston, B., Clabots, C., Roberts, P. L., Billig, M., Riddell, K., Rogers, P., Qin, X., Butler-Wu, S., Price, L. B., Aziz, M., Nicolas-Chanoine, M. H., Debroy, C., Robicsek, A., Hansen, G., Urban, C., Platell, J., Trott, D. J., Zhanel, G., Weissman, S. J., Cookson, B. T., Fang, F. C., Limaye, A. P., Scholes, D., Chattopadhyay, S., Hooper, D. C., and Sokurenko, E. V. (2013). Abrupt emergence of a single dominant multidrug-resistant strain of Escherichia coli. *Journal of Infectious Diseases*, 207(6):919–928.

[74] Kallonen, T., Brodrick, H. J., Harris, S. R., Corander, J., Brown, N. M., Martin, V., Peacock, S. J., and Parkhill, J. (2017). Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Research*, 27(8):1437–1449.

[75] Kantele, A., Kuenzli, E., Dunn, S. J., Dance, D. A. B., Newton, P. N., Davong, V., Mero, S., Pakkanen, S. H., Neumayr, A., Hatz, C., Snaith, A., Kallonen, T., Corander, J., and McNally, A. (2021). Dynamics of intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic setting: a prospective, daily, real-time sampling study. *The Lancet Microbe*, 2(4):e151–e158.

[76] Kantele, A., Lääveri, T., Mero, S., Häkkinen, I. M., Kirveskari, J., Johnston, B. D., and Johnson, J. R. (2020). Despite predominance of uropathogenic/extraintestinal pathotypes among travel-acquired extended-spectrum β-lactamase-producing Escherichia coli, the most commonly associated clinical manifestation is travelers' diarrhea. *Clinical Infectious Diseases*, 70(2):210–2185.

[77] Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004). Pathogenic Escherichia coli. *Nature Reviews Microbiology*, 2(2):123–140.

[78] Katz, L., Griswold, T., Morrison, S., Caravas, J., Zhang, S., Bakker, H., Deng, X., and Carleton, H. (2019). Mashtree: a rapid comparison of whole genome sequence files. *Journal of Open Source Software*, 4(44):1762.

[79] Lasaro, M., Liu, Z., Bishar, R., Kelly, K., Chattopadhyay, S., Paul, S., Sokurenko, E., Zhu, J., and Goulian, M. (2014). Escherichia coli Isolate for Studying Colonization of the Mouse Intestine and Its Application to Two-Component Signaling Knockouts. *Journal of Bacteriology*, 196(9):1723–1732.

[80] Le Gall, T., Clermont, O., Gouriou, S., Picard, B., Nassif, X., Denamur, E., and Tenaillon, O. (2007). Extraintestinal virulence is a coincidental by-product of commensalism in b2 phylogenetic group Escherichia coli strains. *Molecular Biology and Evolution*, 24(11):2373–2384.

[81] Lehtinen, S., Blanquart, F., Croucher, N. J., Turner, P., Lipsitch, M., and Fraser, C. (2017). Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. *Proceedings of the National Academy of Sciences*, 114(5):1075–1080.

[82] Litvak, Y., Byndloss, M. X., Tsolis, R. M., and Bäumler, A. J. (2017). Dysbiotic Proteobacteria expansion: a microbial signature of epithelial dysfunction. *Current Opinion in Microbiology*, 39:1–6.

[83] Long, H., Miller, S. F., Strauss, C., Zhao, C., Cheng, L., Ye, Z., Griffin, K., Te, R., Lee, H., Chen, C. C., and Lynch, M. (2016). Antibiotic treatment enhances the genome-wide mutation rate of target cells. *Proceedings of the National Academy of Sciences of the United States of America*, 113(18):E2498–E2505.

[84] Lopatkin, A. J., Bening, S. C., Manson, A. L., Stokes, J. M., Kohanski, M. A., Badran, A. H., Earl, A. M., Cheney, N. J., Yang, J. H., and Collins, J. J. (2021). Clinically relevant mutations in core metabolic genes confer antibiotic resistance. *Science*, 371(6531).

[85] Lopez, C. A., Rivera-Chávez, F., Byndloss, M. X., and Bäumler, A. J. (2015). The periplasmic nitrate reductase NapABC supports luminal growth of Salmonella enterica serovar typhimurium during colitis. *Infection and Immunity*, 83(9):3470–3478.

[86] Lu, S., Jin, D., Wu, S., Yang, J., Lan, R., Bai, X., Liu, S., Meng, Q., Yuan, X., Zhou, J., Pu, J., Chen, Q., Dai, H., Hu, Y., Xiong, Y., Ye, C., and Xu, J. (2016). Insights into the evolution of pathogenicity of Escherichia coli from genomic analysis of intestinal E. coli of Marmota himalayana in Qinghai–Tibet plateau of China. *Emerging Microbes & Infections*, 5(1):1–9.

[87] Luongo, D., Coppola, A., Treppiccione, L., Bergamo, P., Sorrentino, A., Ferrocino, I., Turroni, S., Neviani, E., Di Cagno, R., Cocolin, L., and Rossi, M. (2017). Modulation of the cytokine profile in Caco-2 cells by faecal lactobacilli and bifidobacteria from individuals with distinct dietary habits. *Cytokine*, 90:80–87.

[88] Lupp, C., Robertson, M. L., Wickham, M. E., Sekirov, I., Champion, O. L., Gaynor, E. C., and Finlay, B. B. (2007). Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host and Microbe*, 2(2):119–129.

[89] 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(3):268–281.

[90] Maltby, R., Leatham-Jensen, M. P., Gibson, T., Cohen, P. S., and Conway, T. (2013). Nutritional Basis for Colonization Resistance by Human Commensal Escherichia coli Strains HS and Nissle 1917 against E. coli O157:H7 in the Mouse Intestine. *PLoS ONE*, 8(1):1–10.

[91] Martinez, J. J., Mulvey, M. A., Schilling, J. D., Pinkner, J. S., and Hultgren, S. J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO Journal*, 19(12):2803–2812.

[92] Mcdaniel, T. K., Jarvis, K. G., Donnenberg, M. S., and Kaper, J. B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 92(5):1664–1668.

[93] McGeachy, M. J., Bak-Jensen, K. S., Chen, Y., Tato, C. M., Blumenschein, W., McClanahan, T., and Cua, D. J. (2007). TGF-β and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain TH-17 cell-mediated pathology. *Nature Immunology*, 8(12):1390–1397.

[94] McNally, A., Cheng, L., Harris, S. R., and Corander, J. (2013). The Evolutionary Path to Extraintestinal Pathogenic, Drug-Resistant Escherichia coli Is Marked by Drastic Reduction in Detectable Recombination within the Core Genome. *Genome Biology and Evolution*, 5(4):699–710.

[95] McNally, A., Kallonen, T., Connor, C., Abudahab, K., Aanensen, D. M., Horner, C., Peacock, S. J., Parkhill, J., Croucher, N. J., and Corander, J. (2019). Diversification of colonization factors in a multidrug-resistant escherichia coli lineage evolving under negative frequency- dependent selection. *mBio*, 10(2):1–19.

[96] McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., Vehkala, M., Välimäki, N., Prentice, M. B., Ashour, A., Avram, O., Pupko, T., Dobrindt, U., Literak, I., Guenther, S., Schaufler, K., Wieler, L. H., Zhiyong, Z., 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(9):1–16.

[97] Melican, K., Sandoval, R. M., Kader, A., Josefsson, L., Tanner, G. A., Molitoris, B. A., and Richter-Dahlfors, A. (2011). Uropathogenic escherichia coli P and type 1 fimbriae act in synergy in a living host to facilitate renal colonization leading to nephron obstruction. *PLoS Pathogens*, 7(2):2–13.

[98] Mohamed, M., Clabots, C., Porter, S. B., Bender, T., Thuras, P., and Johnson, J. R. (2020). Large Fecal Reservoir of Escherichia coli Sequence Type 131-H30 Subclone Strains That Are Shared Within Households and Resemble Clinical ST131-H30 Isolates. *Journal of Infectious Diseases*, 221(10):1659–1668.

[99] Moreno, E., Andreu, A., Pigrau, C., Kuskowski, M. A., Johnson, J. R., and Prats, G. (2008). Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. *Journal of Clinical Microbiology*, 46(8):2529–2534.

[100] Mulvey, M. A., Lopez-Boado, Y. S., Wilson, C. L., Roth, R., Parks, W. C., Heuser, J., and Hultgren, S. J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic Escherichia coli. *Science*, 282(5393):1494–1497.

[101] Nataro, J. P. and Kaper, J. B. (1998). Diarrheagenic Escherichia coli. *Clinical Microbiology Reviews*, 11(1):142–201.

[102] Niccum, B. A., Lee, H., MohammedIsmail, W., Tang, H., and Foster, P. L. (2018). The spectrum of replication errors in the absence of error correction assayed across the whole genome of escherichia coli. *Genetics*, 209(4):1043–1054.

[103] Nicolas-Chanoine, M.-H., Blanco, J., Leflon-Guibout, V., Demarty, R., Alonso, M. P., Canica, M. M., Park, Y.-J., Lavigne, J.-P., Pitout, J., and Johnson, J. R. (2007). Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *Journal of Antimicrobial Chemotherapy*, 61(2):273–281.

[104] 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(PART 4):582–589.

[105] Nielsen, K. L., Stegger, M., Kiil, K., Godfrey, P. A., Feldgarden, M., Lilje, B., Andersen, P. S., and Frimodt-Møller, N. (2017). Whole-genome comparison of urinary pathogenic Escherichia coli and faecal isolates of UTI patients and healthy controls. *International Journal of Medical Microbiology*, 307(8):497–507.

[106] Olesen, B., Frimodt-Møller, J., Leihof, R. F., Struve, C., Johnston, B., Hansen, D. S., Scheutz, F., Krogfelt, K. A., Kuskowski, M. A., Clabots, C., and Johnson, J. R. (2014). Temporal trends in antimicrobial resistance and virulence-associated traits within the Escherichia coli sequence type 131 clonal group and its H30 and H30-Rx Subclones, 1968 to 2012. *Antimicrobial Agents and Chemotherapy*, 58(11):6886–6895.

[107] Olson, A. B., Silverman, M., Boyd, D. A., McGeer, A., Willey, B. M., Pong-Porter, V., Daneman, N., and Mulvey, M. R. (2005). Identification of a progenitor of the CTX-M-9 group of extended-spectrum β-lactamases from Kluyvera georgiana isolated in Guyana. *Antimicrobial Agents and Chemotherapy*, 49(5):2112–2115.

[108] 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(1):1–14.

[109] Östholm-Balkhed, Å., Tärnberg, M., Nilsson, M., Nilsson, L. E., Hanberger, H., and Hällgren, A. (2013). Travel-associated faecal colonization with esbl-producing enterobacteriaceae: Incidence and risk factors. *Journal of Antimicrobial Chemotherapy*, 68(9):2144–2153.

[110] Ouyang, W., Rutz, S., Crellin, N. K., Valdez, P. A., and Hymowitz, S. G. (2011). Regulation and functions of the IL-10 family of cytokines in inflammation and disease. *Annual Review of Immunology*, 29:71–109.

[111] Overdevest, I., Haverkate, M., Veenemans, J., Hendriks, Y., Verhulst, C., Mulders, A., Couprie, W., Bootsma, M., Johnson, J., and Kluytmans, J. (2016). Prolonged colonisation with escherichia coli O25:ST131 versus other extended-spectrum beta-lactamase-producing E. coli in a long-term care facility with high endemic level of rectal colonisation, the Netherlands, 2013 to 2014. *Eurosurveillance*, 21(42).

[112] Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., Fookes, M., Falush, D., Keane, J. A., and Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22):3691–3693.

[113] Paltansing, S., Vlot, J. A., Kraakman, M. E., Mesman, R., Bruijning, M. L., Bernards, A. T., Visser, L. G., and Veldkamp, K. E. (2013). Extended-spectrum β-lactamase-producing enterobacteriaceae among travelers from the Netherlands. *Emerging infectious diseases*, 19(8):1206–1213.

[114] Parlesak, A., Haller, D., Brinz, S., Baeuerlein, A., and Bode, C. (2004). Modulation of cytokine release by differentiated CACO-2 cells in a compartmentalized coculture model with mononuclear leucocytes and nonpathogenic bacteria. *Scandinavian Journal of Immunology*, 60(5):477–485.

[115] Paterson, D. L. and Bonomo, R. A. (2005). Clinical Update Extended-Spectrum Beta-Lactamases : a Clinical Update. *Clinical Microbiology Reviews*, 18(4):657–686.

[116] Peirano, G., Van Der Bij, A. K., Gregson, D. B., and Pitout, J. D. (2012). Molecular epidemiology over an 11-year period (2000 to 2010) of extended-spectrum β-lactamase-producing Escherichia coli causing bacteremia in a centralized Canadian region. *Journal of Clinical Microbiology*, 50(2):294–299.

[117] Perez-Lopez, A., Behnsen, J., Nuccio, S. P., and Raffatellu, M. (2016). Mucosal immunity to pathogenic intestinal bacteria. *Nature Reviews Immunology*, 16(3):135–148.

[118] Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., and Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature*, 409(6819):529–33.

[119] Petty, N. K., Ben Zakour, N. L., Stanton-Cook, M., Skippington, E., Totsika, M., Forde, B. M., Phan, M.-D., Gomes Moriel, D., Peters, K. M., Davies, M., Rogers, B. A., Dougan, G., Rodriguez-Bano, J., Pascual, A., Pitout, J. D. D., Upton, M., Paterson, D. L., Walsh, T. R., Schembri, M. A., and Beatson, S. A. (2014). Global dissemination of a multidrug resistant Escherichia coli clone. *Proceedings of the National Academy of Sciences*, 111(15):5694–5699.

[120] Pires, J., Kraemer, J. G., Kuenzli, E., Kasraian, S., Tinguely, R., Hatz, C., Endimiani, A., and Hilty, M. (2019). Gut microbiota dynamics in travelers returning from India colonized with extended-spectrum cephalosporin-resistant Enterobacteriaceae: A longitudinal study. *Travel Medicine and Infectious Disease*, 27(May 2018):72–80.

[121] Poirel, L., Lartigue, M.-F., Decousser, J.-w., and Nordmann, P. (2005). ISEcp1B-Mediated Transposition of blaCTX-M in Escherichia coli. *Antimicrobial Agents and Chemotherapy*, 49(1):447–450.

[122] Porat, R., Clark, B. D., Wolff, S. M., and Dinarello, C. A. (1991). Enhancement of growth of virulent strains of Escherichia coli by interleukin-1. *Science*, 254(5030):430–432.

[123] Price, L. B., Johnson, J. R., Aziz, M., Clabots, C., Johnston, B., Tchesnokova, V., Nordstrom, L., Billig, M., Chattopadhyay, S., Stegger, M., Andersen, P. S., Pearson, T., Riddell, K., Rogers, P., Scholes, D., Kahl, B., Keim, P., and Sokurenko, E. V. (2013). The epidemic of extended-spectrum-β-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. *mBio*, 4(6):1–10.

[124] Rasko, D. A., Webster, D. R., Sahl, J. W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E. E., Sebra, R., Chin, C.-S., Iliopoulos, D., Klammer, A., Peluso, P., Lee, L., Kislyuk, A. O., Bullard, J., Kasarskis, A., Wang, S., Eid, J., Rank, D., Redman, J. C., Steyert, S. R., Frimodt-Møller, J., Struve, C., Petersen, A. M., Krogfelt, K. A., Nataro, J. P., Schadt, E. E., and Waldor, M. K. (2011). Origins of the E. coli Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany. *New England Journal of Medicine*, 365(8):709–717.

[125] Rivera-Chávez, F. and Bäumler, A. J. (2015). The Pyromaniac Inside You: Salmonella Metabolism in the Host Gut. *Annual Review of Microbiology*, 69(1):31–48.

[126] Rodríguez, M. M., Power, P., Radice, M., Vay, C., Famiglietti, A., Galleni, M., Ayala, J. A., and Gutkind, G. (2004). Chromosome-encoded CTX-M-3 from Kluyvera ascorbata: A possible origin of plasmid-borne CTX-M-1-derived cefotaximases. *Antimicrobial Agents and Chemotherapy*, 48(12):4895–4897.

[127] Romero, L., López, L., Rodríguez-Baño, J., Hernández, J. R., Martínez-Martínez, L., and Pascual, A. (2005). Long-term study of the frequency of Escherichia coli and Klebsiella pneumoniae isolates producing extended-spectrum β-lactamases. *Clinical Microbiology and Infection*, 11(8):625–631.

[128] Ruppé, E., Armand-Lefèvre, L., Estellat, C., Consigny, P. H., El Mniai, A., Boussadia, Y., Goujon, C., Ralaimazava, P., Campa, P., Girard, P. M., Wyplosz, B., Vittecoq, D., Bouchaud, O., Le Loup, G., Pialoux, G., Perrier, M., Wieder, I., Moussa, N., Esposito-Farèse, M., Hoffmann, I., Coignard, B., Lucet, J. C., Andremont, A., and Matheron, S. (2015). High Rate of Acquisition but Short Duration of Carriage of Multidrug-Resistant Enterobacteriaceae after Travel to the Tropics. *Clinical Infectious Diseases*, 61(4):593–600.

[129] Russo, T. A. and Johnson, J. R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of Escherichia coli: ExPEC. *Journal of Infectious Diseases*, 181(5):1753–1754.

[130] Sahl, J. W., Matalka, M. N., and Rasko, D. A. (2012). Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments. *Applied and Environmental Microbiology*, 78(14):4884–4892.

[131] Sansonetti, P. J., Hale, T. L., Dammin, G. J., Kapfer, C., Collins, H. H., and Formal, S. B. (1983). Alterations in the pathogenicity of Escherichia coli K-12 after transfer of plasmid and chromosomal genes from Shigella flexneri. *Infection and Immunity*, 39(3):1392–1402.

[132] Santos, R. L., Raffatellu, M., Bevins, C. L., Adams, L. G., Tükel, Ç., Tsolis, R. M., and Bäumler, A. J. (2009). Life in the inflamed intestine, Salmonella style. *Trends in Microbiology*, 17(11):498–506.

[133] Sarkar, S., Hutton, M. L., Vagenas, D., Ruter, R., Schüller, S., Lyras, D., Schembri, M. A., and Totsika, M. (2018). Intestinal colonization traits of pandemic multidrug-resistant Escherichia coli ST131. *Journal of Infectious Diseases*, 218(6):979–990.

[134] Sarokin, L. and Carlson, M. (1984). Nucleic Acids Research Nucleic Acids Research. *Methods*, 12(21):8235–8251.

[135] Sauer, M. M., Jakob, R. P., Eras, J., Baday, S., Eriş, D., Navarra, G., Bernèche, S., Ernst, B., Maier, T., and Glockshuber, R. (2016). Catch-bond mechanism of the bacterial adhesin FimH. *Nature Communications*, 7.

[136] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682.

[137] Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069.

[138] Simonet, V., Malléa, M., and Pagès, J. M. (2000). Substitutions in the eyelet region disrupt cefepime diffusion through the Escherichia coli OmpF channel. *Antimicrobial Agents and Chemotherapy*, 44(2):311–315.

[139] Sonnenborn, U., Schmidt, M. A., Sabharwal, H., and Cichon, C. (2016). Human Intestinal Epithelial T84 and Monocytic THP-1 Cells after Apical or Basolateral Infection. *Infection and Immunity*, 84(9):2482–2492.

[140] Soysal, N., Mariani-Kurkdjian, P., Smail, Y., Liguori, S., Gouali, M., Loukiadis, E., Fach, P., Bruyand, M., Blanco, J., Bidet, P., and Bonacorsi, S. (2016). Enterohemorrhagic Escherichia coli hybrid pathotype O80:H2 as a new therapeutic challenge. *Emerging Infectious Diseases*, 22(9):1604–1612.

[141] Spees, A. M., Wangdi, T., Lopez, C. A., Kingsbury, D. D., Xavier, M. N., Winter, S. E., Tsolis, R. M., and Bäumler, A. J. (2013). Streptomycin-Induced Inflammation Enhances Escherichia coli Gut Colonization Through Nitrate Respiration. *mBio*, 4(4):1–10.

[142] Sprouffske, K. and Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1):17–20.

[143] Stoesser, N., Sheppard, A. E., Pankhurst, L., de Maio, N., Moore, C. E., Sebra, R., Turner, P., Anson, L. W., Kasarskis, A., Batty, E. M., Kos, V., Wilson, D. J., Phetsouvanh, R., Wyllie, D., Sokurenko, E., Manges, A. R., Johnson, T. J., Price, L. B., Peto, T. E., Johnson, J. R., Didelot, X., Sarah Walker, A., and Crook, D. W. (2016). Evolutionary history of the global emergence of the Escherichia coli epidemic clone ST131. *mBio*, 7(2):1–15.

[144] Tängdén, T., Cars, O., Melhus, Å., and Löwdin, E. (2010). Foreign travel is a major risk factor for colonization with Escherichia coli producing CTX-M-type extended-spectrum β-lactamases: A prospective study with Swedish volunteers. *Antimicrobial Agents and Chemotherapy*, 54(9):3564–3568.

[145] Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal Escherichia coli. *Nature Reviews Microbiology*, 8(3):207–217.

[146] Teunis, P., Takumi, K., and Shinagawa, K. (2004). Dose Response for Infection by Escherichia coli O157:H7 from Outbreak Data. *Risk Analysis*, 24(2):401–407.

[147] Thiennimitr, P., Winter, S. E., Winter, M. G., Xavier, M. N., Tolstikov, V., Huseby, D. L., Sterzenbach, T., Tsolis, R. M., Roth, J. R., and Baumler, A. J. (2011). Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proceedings of the National Academy of Sciences*, 108(42):17480–17485.

[148] Thorpe, H. A., Bayliss, S. C., Sheppard, S. K., and Feil, E. J. (2018). Piggy: A rapid, large-scale pan-genome analysis tool for intergenic regions in bacteria. *GigaScience*, 7(4):1–11.

[149] Totsika, M., Beatson, S. A., Sarkar, S., Phan, M. D., Petty, N. K., Bachmann, N., Szubert, M., Sidjabat, H. E., Paterson, D. L., Upton, M., and Schembri, M. A. (2011). Insights into a multidrug resistant Escherichia coli pathogen of the globally disseminated ST131 lineage: Genome analysis and virulence mechanisms. *PLoS ONE*, 6(10).

[150] Van De Walle, J., Hendrickx, A., Romier, B., Larondelle, Y., and Schneider, Y. J. (2010). Inflammatory parameters in Caco-2 cells: Effect of stimuli nature, concentration, combination and cell differentiation. *Toxicology in Vitro*, 24(5):1441–1449.

[151] van der Donk, C. F., van de Bovenkamp, J. H., de Brauwer, E. I., de Mol, P., Feldhoff, K. H., Kalka-Moll, W. M., Nys, S., Thoelen, I., Trienekens, T. A., and Stobberingh, E. E. (2012). Antimicrobial Resistance and Spread of Multi Drug Resistant Escherichia coli Isolates Collected from Nine Urology Services in the Euregion Meuse-Rhine. *PLoS ONE*, 7(10):3–7.

[152] Vejborg, R. M., Friis, C., Hancock, V., Schembri, M. A., and Klemm, P. (2010). A virulent parent with probiotic progeny: Comparative genomics of Escherichia coli strains CFT073, Nissle 1917 and ABU 83972. *Molecular Genetics and Genomics*, 283(5):469–484.

[153] Vidal, S., Bredin, J., Pagès, J. M., and Barbe, J. (2005). β-lactam screening by specific residues of the OmpF eyelet. *Journal of Medicinal Chemistry*, 48(5):1395–1400.

[154] Viladomiu, M., Metz, M. L., Lima, S. F., Jin, W.-b., Chou, L., Guo, C.-J., Diehl, G. E., Simpson, K. W., Scherl, E. J., and Longman, R. S. (2021). Adherent-invasive E. coli metabolism of propanediol in Crohn's disease regulates phagocytes to drive intestinal inflammation. *Cell Host & Microbe*, pages 1–13.

[155] Vimont, S., Boyd, A., Bleibtreu, A., Bens, M., Goujon, J. M., Garry, L., Clermont, O., Denamur, E., Arlet, G., and Vandewalle, A. (2012). The CTX-M-15-Producing Escherichia coli Clone O25b: H4-ST131 Has High Intestine Colonization and Urinary Tract Infection Abilities. *PLoS ONE*, 7(9).

[156] Von Mentzer, A., Connor, T. R., Wieler, L. H., Semmler, T., Iguchi, A., Thomson, N. R., Rasko, D. A., Joffre, E., Corander, J., Pickard, D., Wiklund, G., Svennerholm, A. M., Sjöling, Å., and Dougan, G. (2014). Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. *Nature Genetics*, 46(12):1321–1326.

[157] Walk, S. T. (2015). The "Cryptic" Escherichia. *EcoSal Plus*, 6(2):ecosalplus.ESP–0002–2015.

[158] Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L., Donnenberg, M. S., and Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):17020–17024.

[159] Williams, K. P. (2003). Traffic at the tmRNA gene. *Journal of Bacteriology*, 185(3):1059–1070.

[160] Winter, S. E., Lopez, C. A., and Bäumler, A. J. (2013a). The dynamics of gut-associated microbial communities during inflammation. *EMBO Reports*, 14(4):319–327.

[161] Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., Roth, J. R., and Bäumler, A. J. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, 467(7314):426–429.

[162] Winter, S. E., Winter, M. G., Xavier, M. N., Thiennimitr, P., Poon, V., Keestra, A. M., Laughlin, R. C., Gomez, G., Wu, J., Lawhon, S. D., Popova, I. E., Parikh, S. J., Adams, L. G., Tsolis, R. M., Stewart, V. J., and Bäumler, A. J. (2013b). Host-derived nitrate boosts growth of E. coli in the inflamed gut. *Science*, 339(6120):708–711.

[163] Woerther, P. L., Andremont, A., and Kantele, A. (2017). Travel-acquired ESBL-producing Enterobacteriaceae: impact of colonization at individual and community level. *Journal of travel medicine*, 24(1):S29–S34.

[164] Woerther, P. L., Burdet, C., Chachaty, E., and Andremont, A. (2013). Trends in human fecal carriage of extended-spectrum β-lactamases in the community: Toward the globalization of CTX-M. *Clinical Microbiology Reviews*, 26(4):744–758.

[165] Wotzka, S. Y., Kreuzer, M., Maier, L., Arnoldini, M., Nguyen, B. D., Brachmann, A. O., Berthold, D. L., Zünd, M., Hausmann, A., Bakkeren, E., Hoces, D., Gül, E., Beutler, M., Dolowschiak, T., Zimmermann, M., Fuhrer, T., Moor, K., Sauer, U., Typas, A., Piel, J., Diard, M., Macpherson, A. J., Stecher, B., Sunagawa, S., Slack, E., and Hardt, W. D. (2019). Escherichia coli limits Salmonella Typhimurium infections after diet shifts and fat-mediated microbiota perturbation in mice. *Nature Microbiology*, 4(12):2164–2174.

[166] Wright, K. J., Seed, P. C., and Hultgren, S. J. (2007). Development of intracellular bacterial communities of uropathogenic Escherichia coli depends on type 1 pili. *Cellular Microbiology*, 9(9):2230–2241.

[167] Yen, D., Cheung, J., Scheerens, H., Poulet, F., McClanahan, T., Mckenzie, B., Kleinschek, M. A., Owyang, A., Mattson, J., Blumenschein, W., Murphy, E., Sathe, M., Cua, D. J., Kastelein, R. A., and Rennick, D. (2006). IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *Journal of Clinical Investigation*, 116(5):1310–1316.

[168] Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., and Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, 67(11):2640–2644.

[169] Zhang, R., Liu, L., Zhou, H., Chan, E. W., Li, J., Fang, Y., Li, Y., Liao, K., and Chen, S. (2017). Nationwide Surveillance of Clinical Carbapenem-resistant Enterobacteriaceae (CRE) Strains in China. *EBioMedicine*, 19:98–106.

[170] Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, 39(SUPPL. 2):347–352.

[171] 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(1):138–152.

[172] Zong, Z., Fenn, S., Connor, C., Feng, Y., and McNally, A. (2018). Complete genomic characterization of two Escherichia coli lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital. *Journal of Antimicrobial Chemotherapy*, 73(9):2340–2346.

Appendix A

Supplementary Data

A.1 Fluorescent Plasmid Fitness Cost

Transforming bacterial strains has the potential to affect their growth kinetics, specifically plasmids may impose a fitness cost on their host reducing their growth rate. In addition the plasmids used in Chapter 4 were not identical so that bacterial strains could be distinguished based on fluorescence. It is therefore possible that they exert different fitness costs on the transformed strains. In order to address this the plasmid transformed strains were compared to their parental counterparts to measure if they impose any fitness cost or alteration to growth kinetics. The fitness costs were assessed in nutrient rich LB broth and minimal DMEM media, strains were adjusted to equivalent optical densities before inoculating media. Both K12 and F104 showed a moderate delay in growth when transformed with plasmids and grown in LB (Figure A.1). There was minimal growth of K12 in DMEM making inferences about plasmid fitness costs difficult. The F104 strain displayed a reduced rate of growth when transformed with the plasmid and grown in DMEM. The strain 811E7 displayed a minor delay in growth in both LB and DMEM when the tagging plasmid was present. Whilst F016 displayed a moderate delay in growth in LB that was heightened when grown in DMEM. A logistic regression model was fitted to the data in order to calculate generation, which

were not significantly altered for any of the strains assayed under either media condition (Figure A.2a & b). While the strains do not display a significant alteration to generation time there is a delay in exiting the lag phase when the plasmid is present. To investigate this the 'time to half carrying capacity' was calculated from the logistic regression model (Figure A.2c & d). In LB all strains assayed displayed a significant increase in the time to half-carrying capacity when transformed with a plasmid. Whilst in DMEM there was no significant difference between parental and transformed strains.

In addition to the growth kinetics, direct competition between parental and plasmid transformed strains was compared. Strains were mixed in a 1 to 1 ratio and inoculated into fresh media with growth after 24 hours being assessed. Total growth was measured by CFU assay on non-selective agar, while plasmid transformants were distinguished with selective agar. The strains 811E7 (ST10) and F016 (ST131) were assayed in both LB and DMEM. In both strains and under both conditions the plasmid transformant was out-competed by its parental counterpart (Figure A.3). This suggests that the presence of either fluorescent plasmid impairs the growth of the transformed strain.

The plasmids used in this work (pFPV25.1 and pCherry8) are both small plasmids encoding fluorescent reporter proteins. Neither plasmid encodes plasmid maintenance systems such as partionining machinery to ensure that the plasmid is maintained in daughter cells. It is therefore possible that plasmids are being lost in rapidly dividing populations. The assays presented cannot distinguish between parental cells and transformed cells that have lost the plasmid.

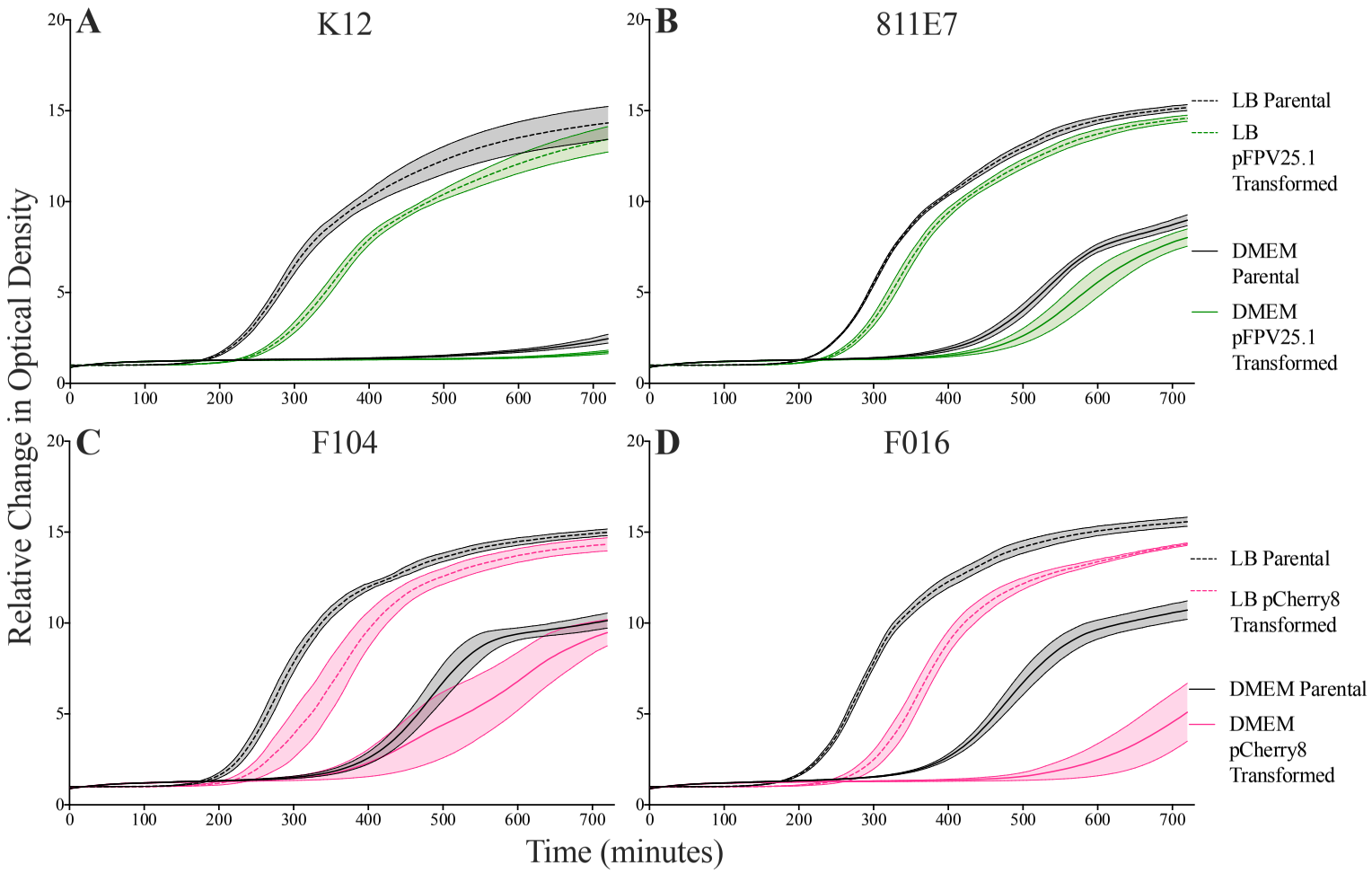

Fig. A.1 Growth dynamics of strains transformed with plasmids and their parental counterparts. Parental strains (black lines) and their transformants (green lines for GFP, pink lines for mCherry) were cultured in LB (dashed line) and DMEM (solid line). The strains K12 (A), 811E7 (B), F104 (C) and F016 (D) were assayed.

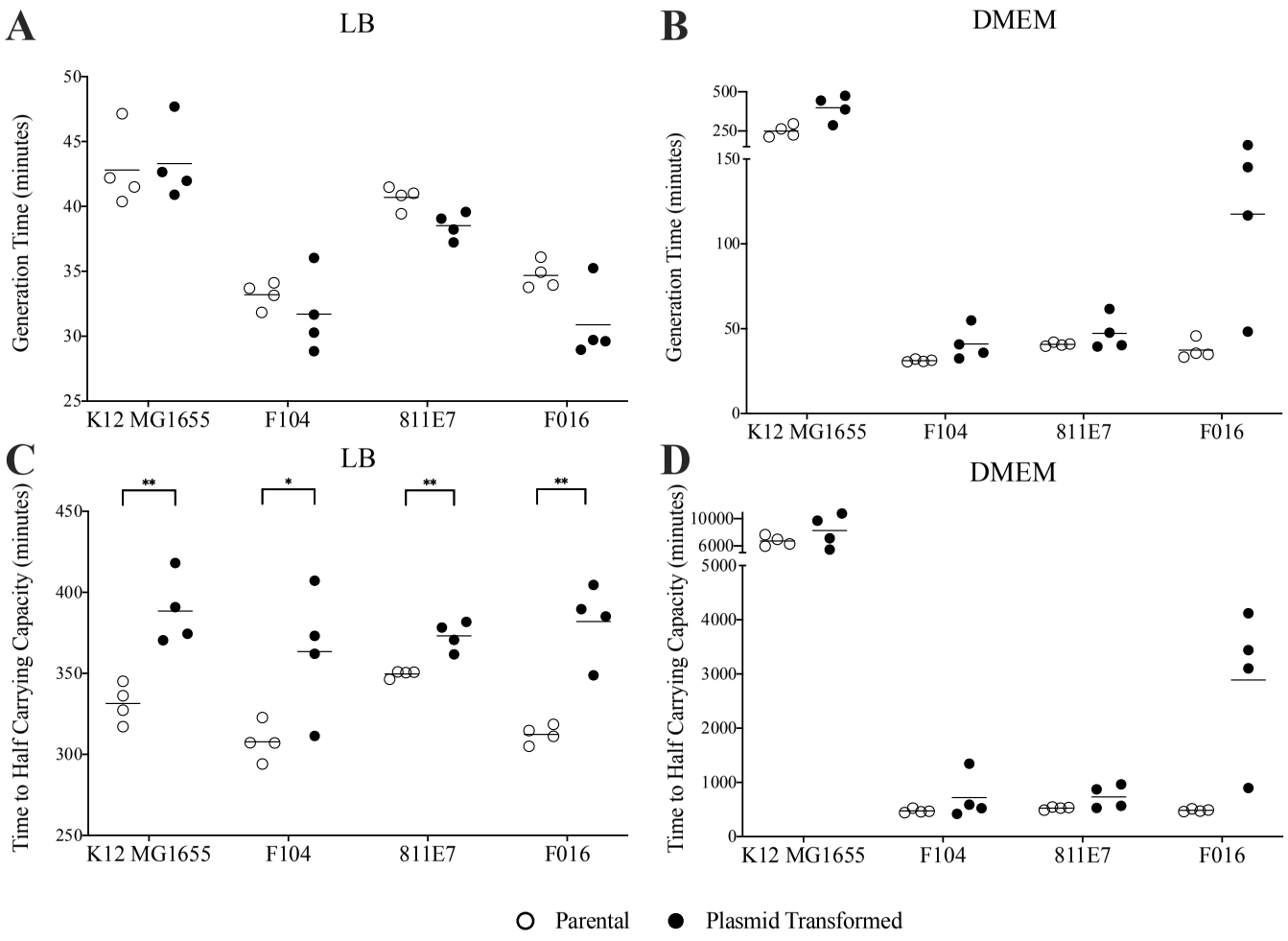

Fig. A.2 Metrics from growth curve dynamics. A logistic regression model was fitted to the growth data from Figure A.1 using the R package GrowthCurver [142]. The calculated generation time (A and B) as well as the time to reach half carrying capacity (C and D) are plotted for the parental (open circles) and plasmid transformants (closed circles). Lines indicate mean, n=4. Significance determined by t-test, * $p<0.05$, ** $p<0.01$.

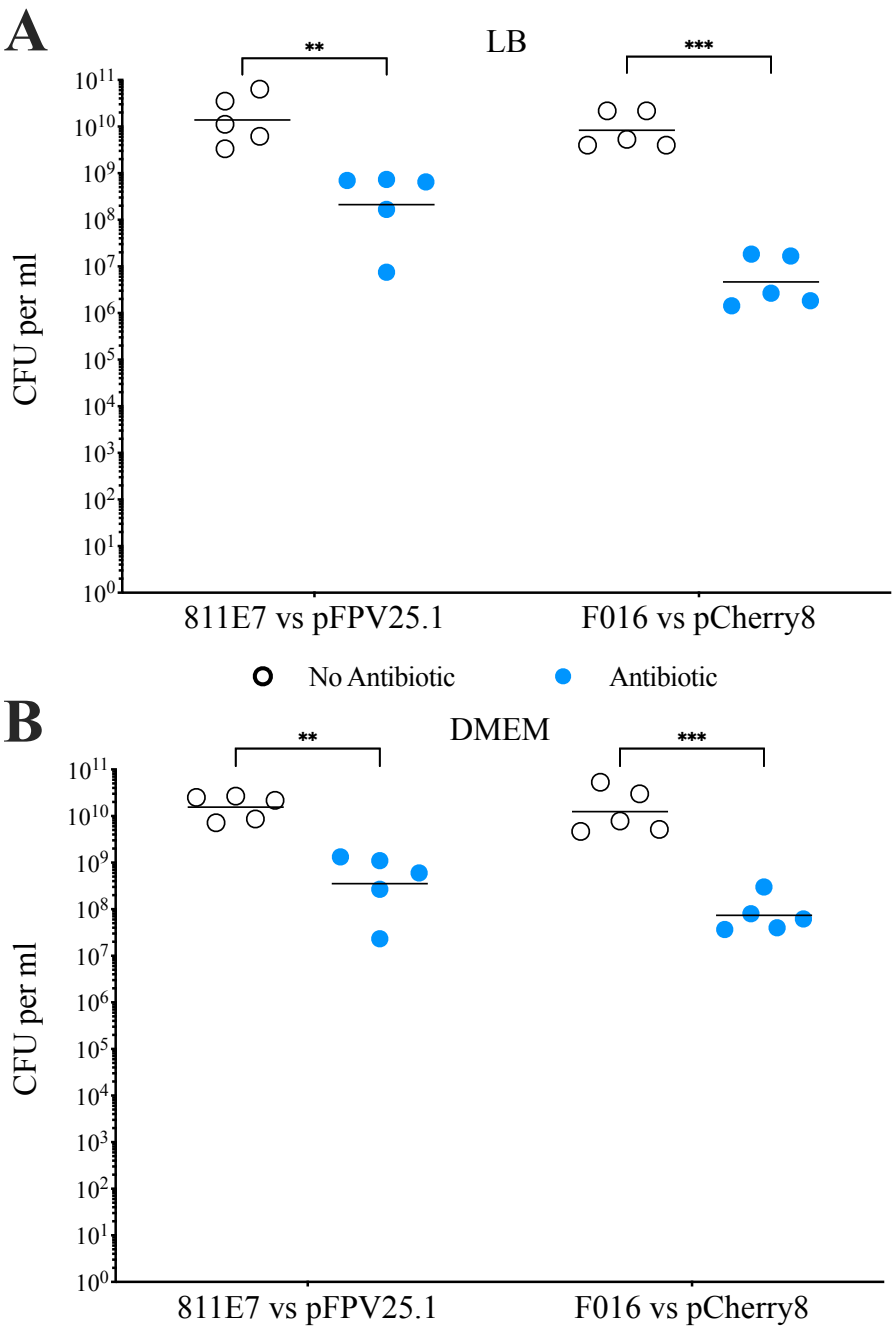

Fig. A.3 Competition assay between parental and plasmid transformed strains. Parental and transformed strains were inoculated at equivalent densities into the same volume of LB or DMEM media and grown for 24 hours. Total growth was enumerated by plating onto non-selective agar (open cirlces) while plasmid carrying strains were identified by plating onto selective agar (blue dot). Lines indicate geometric mean. Significance determined by t-test on log normalised CFU values, ** $p<0.01$, *** $p<0.001$.