Why Do Some Human Associated *Escherichia coli* Strains Lack Antibiotic Resistance?

Veronica Chineye Alozie-chidi

February 2022

A thesis submitted for the degree of Doctor of Philosophy at
The Australian National University

Declaration

I declare that this thesis titled ‘Why do some human associated *E. coli* strains lack antibiotic resistance’ is the product of my original work. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Veronica Chineye Alozie-chidi

February 2022

Acknowledgements

I am enormously grateful to my chair supervisor Professor David Michael Gordon. Thank you for your objective evaluation of the research at every stage and constructive criticism of the work and write-ups. All your efforts are greatly appreciated. I am very grateful to members of my supervisory panel, Professor Paul Pavli and Dr. Claire O Brien for their constructive feedback and continuous support and guidance throughout my PhD program.

I would like to thank the Ecology and Evolution heads, Professor Craig, and Michael for providing an enabling environment for this program. Thank you so much the HDR convenor, Professor Celeste, for your encouraging words. I am so grateful.

I thank the technical team, Jack, Wes and the administrator, Audra, for always ready to help when call upon. I also appreciate the RSB IT team, they are always there when I run into any challenge with any of my installed application tools.

I also appreciate the academic skills unit, the coaches that went through my drafts, Nyree Mason, who patiently taught me on how to go about my statistical analysis, and Candida Spence, who kindly showed me how to arrange and compile my chapters. I am very grateful.

I appreciate all my lab / office mates, Mah, Rasel, Judith, Truc, Angel, Buddi. Thank you, guys. It was wonderful knowing you all.

My husband and children have been very patient and understanding throughout this period. I am grateful that I have them and I appreciate them.

This research is supported by an Australian Government Research Training Program (AGRTP) Scholarship. I thank The Australian national University and Australia for giving me this great opportunity.

Thank you.

Abstract

E. coli ST lineages 131, 73 and 95 of phylogroup B2 are globally pathogenic, and are proportionally distributed in most hospital and community acquired human extra-intestinal diseases and in the gut. However, several pieces of evidence have shown that ST131 is highly multi-drug resistance while STs 73 and 95 are largely susceptible to antibiotics. Limited information exists to explain this. To investigate this, differences in the genomic diversity of plasmids in the lineages were characterised using bioinformatics tools. Plasmid-host dynamics were investigated by examining rate of plasmid transfers using plasmid transfer experiments. To assess the impacts of the plasmids on transfer rates, genome wide association studies were carried out, and to determine how well the plasmids are maintained within the host, stability experiments were done. Finally, fitness costs of the plasmid's carriage within the host over time using competition experiments and the evolutionary history existing among the two common F-plasmid sequence types were examined and compared using genome alignments techniques and phylogenetic analysis. ST73 lineage was observed to carry more of mobilizable plasmids compared to the other lineages. IncFII and IncFIB plasmids were common in the *E. coli* strains analysed. The presence of different classes of antimicrobial resistance genes and frequent occurrence of fluoroquinolone resistance determinant were observed in ST131. Most ST131 plasmids are associated with O25H4:*fimH30*, and the latter are adapted to multiple resistance genes and integrons, mostly detected in F-plasmid sequence type F29:A-:B10. ST95 and ST73 lineages also share this F-plasmid sequence type but most strains are associated with F51:A-:B10, which is also dominated by integrons, but fewer multiple antimicrobial genes. These lineages also encode many plasmid encoded virulence genes and bacteriocins. No diversity was observed in the rate of plasmid transfers among the ST

lineages, however there were some variations between the F-plasmid sequence types and rates of plasmid transfer. This association was based on the presence of some variable genes: *traC*, *traF*, *traH*, *traR* found in the two plasmid types. The presence of the multiple *tra* genes positively enhanced the rates of plasmid transfer. Besides, the conjugal transfer, F-plasmid sequence type effects were also observed in plasmid stability frequencies and host fitness in all strains. Gene content of the two common F-plasmid sequence types revealed that the plasmids harboured about 74% of genes involved in IncF plasmid conjugative transfer. Other genes detected in these sequence types were mostly involved in plasmid stabilization, partition and regulation, regulators of plasmid encoded virulence genes, programmed cell death and operon control. The current study suggests that the success of plasmid-bacterium associations in ST73 and ST95 that made them globally pathogenic, even though antibiotic susceptible can be attributed majorly on the genetic content of the plasmids they encode, especially the F-plasmid sequence types, F51:A-:B10 and F29:A-:B10. These F-plasmid sequence types encode advantageous genes, are highly pathogenic, stable and exert minimal effects on the bacterial host that harbour them. This study has therefore demonstrated that plasmids associated with specific F-plasmid sequence type displayed genotypic characteristics indicative of adaptation, pathogenesis, and persistence in ExPEC diseases. The distinct functions of the *tra* genes encoded by these plasmids in conjugal transfer is pivotal in the adaptation, evolution, and survival of *E. coli* lineages and finally, the effects of F-plasmid sequence type on plasmid stability and host fitness play important roles on the ability of the plasmids to associate with new bacterial hosts and consequently on the evolution of plasmid-mediated antibiotic resistance.

Table of Contents

List of Tables

List of Figures

Chapter 1 General Introduction

1.1 Antibiotic Resistance

Antibiotics are drugs used to prevent and cure infections in humans, animals and even plants. Antibiotic resistance (AR) occurs when microbes evolve mechanisms that protect them from the adverse effects of antibiotics. According to World Health Organization, AR has become a threat to world health and economic growth. Serious illnesses, untreatable infections, increased risks of out-of-control diseases, extended hospital stays, rise in medical bills and increased mortality are some of the challenges this poses (AMR; WHO, 2020). Proliferation of bacteria that are resistant to drugs because of their capacity to acquire new resistance genes is a threat to the ability to treat everyday diseases. The global rise and spread of multi-resistant bacteria has further increased the incidences of these disease outbreaks. These group of bacteria have become resistant to at least three or more antimicrobials that were previously used to treat them (AMR; WHO, 2020). Antibiotic resistant bacteria are known to be the causal agents of frequently encountered bacterial infections, such as urinary tract infections, diarrhoea, bacteraemia, neonatal meningitis, pneumonia, as well as uncommon infections such as disease outbreaks that are specific to patients in intensive-care units (AMR; WHO, 2020; Fratamico *et al*., 2016). These diseases are frequently caused by *Escherichia coli (E. coli)*, which are associated with high levels of antimicrobial resistance (AMR) and may cause intestinal or extra-intestinal infections (Jang *et al.*, 2017). *E. coli* may be acquired from either community or hospital sources (Horner *et al*., 2014; Brisse *et al*., 2012). Urinary tract infections (UTI) caused by *E. coli* are frequently treated with fluoroquinolone antibiotics, however fluoroquinolone resistance is now common among *E. coli* according to the Global Antimicrobial

Resistance and Surveillance system (GLASS) report (AMR; WHO, 2020). High rates of resistance to ciprofloxacin, another antibiotic used to treat UTIs caused by *E. coli*, have also been observed globally, ranging from 8.4% to 92.9% for *E. coli* and 4.1% to 79.4% for *Klebsiella pneumonia*, another member of the Enterobacteriaceae that commonly cause UTI. Resistance to a last resort antibiotic used to treat carbapenem resistant Enterobacteriaceae (CRE), colistin, has been observed in multiple countries and is of concern due to lack of other options for treating such infections. Of additional concern is the high rates of resistance of *E. coli* to third generation cephalosporins (3GC), which are used to treat bloodstream infections. The 2019 GLASS report found that 36% of *E.coli* causing bloodstream infections, surveyed across 49 countries, were resistant to 3GC.

1.2 *Escherichia coli*

E. coli belongs to the family Enterobacteriaceae, a large family of bacteria that is represented by a diverse group of rod shaped, Gram-negative bacteria. *E. coli* naturally live as commensals in the intestinal tract of humans and animals. The *E. coli* pangenome (the full set of genes from all species within a clade) consists of around 76,000 gene families (Touchon *et al*., 2020). The average *E. coli* genome is about 5Mb in size and consists of about 4700 genes. A pangenome is made up of core genes, which are genes present in all strains of a species, and variable genes, which are genes not present in all strains and are dispensable. The core genome of *E. coli* consists of about 2000 genes (Touchon *et al*., 2009).

Multi-locus sequence typing (MLST) is an established typing technique often used for *E. coli.* Three distinct multi-locus sequence typing (MLST) have been identified: the Mark Achtman scheme that uses a set of 7 housekeeping genes (Wirth *et al*., 2006), the Pasteur institute scheme that uses 8 housekeeping genes (Jaureguy *et al*., 2008) and the Thomas Whittam scheme that uses up to 15 housekeeping genes (Qi *et al*., 2004). The genes used by these three MLST schemes are mostly different. The most widely used scheme is Mark Achtman's MLST method. It is a technique in which *E. coli* lineages are allocated a sequence type (ST) with a numerical label. A unique gene sequence is known as an allele, and it is numbered. A unique combination of the alleles for the seven housekeeping genes make up a sequence type (ST) (Manges *et al*., 2015).

E. coli is subdivided into phylogroups whose major representatives are phylogroups A, B1, B2 and D. Phylogroup B2 typically consists of three major sequence types (STs): ST131, ST95 and ST73. These STs and ST69 from phylogroup D account for majority of human associated *E. coli* infections (Riley, 2014). Most strains *of E. coli* are harmless inhabitants of the gut of warm-blooded animals (Kaper *et al.*, 2004), however some strains become pathogenic when opportunities arise causing various diseases.

The pathogenic strains of *E. coli* consist of the intestinal and extra-intestinal pathogenic *E. coli,* and they are responsible for a wide range of highly life-threatening diseases (Jang *et al.*, 2017). Intestinal pathogenic *E. coli* groups have been well studied and are classified according to their virulence factors. They are known to colonise regions within the intestinal tract, where they cause several diseases. Common pathovars that cause these diseases include enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli*

(EIEC), enterotoxigenic *E. coli* (ETEC), enterohaemorrhagic *E. coli* (EHEC) and enteroaggregative *E. coli* (EAEC) (Jang *et al*., 2017). The EPEC strains have been implicated as the source of dysentery amongst children in third world countries. However, EHEC *E. coli* strains are a major cause of intestinal infection in the developed world (Riley, 2014). The EHEC *E. coli* is a shiga toxin-producing *E. coli* that has been known to cause haemolytic uraemia syndrome and haemorrhagic colitis, a major cause of acute renal failure that often leads to death (Kaper *et al*., 2004). Extra-intestinal pathogenic *E. coli* strains also known as ExPEC are responsible for a great number of varying extra-intestinal diseases (Johnson *et al*., 2006; Russo and Johnson, 2003). They colonise sites outside the intestinal system causing diseases, such as cystitis, bacteraemia, pneumonia, neonatal meningitis (Fratamico *et al*., 2016; Gordon, 2010; Smith *et al*., 2007). These diseases are usually acquired from hospitals or communities (Dale and Woodford, 2015).

1.2.1 *Escherichia coli* ST131

ST131 is currently the most researched ExPEC lineage amidst the extra-intestinal pathogenic *E. coli* that have been isolated globally. This lineage was initially isolated in the 2000s. Since then, it has successfully spread all over the world, perhaps because of its tendency to be multidrug-resistant (Nicolas-Chanoine *et al*., 2014).

Strains belonging to ST131 have one of the following major serotypes: O16:H5, O25:H4 and NT: H4 (Olesen *et al*., 2013; Dahbi *et al*., 2013; Matsumura *et al*., 2012). The O25:H4 serotype is known for the acquisition of fluoroquinolone resistance. In addition, seven *fimH* types have been identified in this lineage, with *fimH30* being the

most frequent, followed by *fimH22, fimH35* and *fimH41* (Johnson *et al*., 2013). Most *fimH* types are linked to particular serotypes: for example, majority of O25:H4 serotypes are positive for *fimH30* while O16:H5 serotype is associated with *fimH4*1.

Most ST131 strains have acquired resistance to all classes of antibiotics used to cure infections of humans and animals. Isolates of ST131 were more specifically resistant to amikacin, quinolones / fluoroquinolones, amoxicillin-clavulanic acid, ciprofloxacin, ampicillin / amoxicillin and piperacillin-tazobactam, than non-ST131 isolates (Skov and Monnet, 2016; Liu *et al*., 2016; Zhang *et al*., 2014; Chen *et al*., 2014). These drugs are largely used to cure diseases caused by *E. coli* acquired from community sources. Variety of virulence factor genes are also commonly detected in ST131 strains. They include serum resistance associated gene (*traT*), aerobactin (*iucD*), yersiniabactin receptor (*fyuA*), group 2 capsule synthesis (*kpsMII*), pathogenicity island marker (*malX*), outer membrane receptor (*ompT*), type 1 fimbriae (*fimH*), aerobactin receptor (*iutA*), uropathogen-specific gene (*usp*), adhesin siderophore receptor (*iha*) and secreted autotransporter toxin (*sat*) (Blanco *et al*., 2013; van der Bij *et al*., 2012; Johnson *et al*., 2010; Nicolas-Chanoine *et al*., 2008). It has also been reported that ST131 strains that are resistant to quinolones and fluoroquinolones possess fewer virulence genes than their antibiotic susceptible counterparts (Moreno *et al*., 2006).

Apart from extra-intestinal sites, ST131 strains have also been encountered in the stools of symptomless individuals (Escobar-Paramo *et al*., 2006; Escobar-Paramo *et al*., 2004). Multi-drug resistant (MDR) strains of ST131, isolated from clinical sources, commonly exhibit fluoroquinolone resistance, due to the presence of specific gene mutations in *par and gyr* (Nicolas-Chanoine *et al*., 2008). The presence of encoded

mobile element *CTX-M-15,* in some of the strains of this lineage has also been reported (Price *et al*., 2013). Despite CTX-M-15 being a mobile genetic element, clonal expansion is largely responsible for the proliferation of fluoroquinolone resistance and CTX-M-15, a group of enzymes in a family of class A extended-spectrum β-lactamases (ESBLs) (Nicolas-Chanoine *et al*., 2014; Price *et al*., 2013; Coque *et al*., 2008; Nicolas-Chanoine *et al*., 2008).

1.2.2 *Escherichia coli* ST95

The ST95 lineage is a global and genetically diverse human-associated *E. coli* lineage. Several serotypes have been observed in this lineage, including O18:K1:H7, O1:K1:H7, O2:K1:H7 O2a:H4, O2a:H7, O45a:H7, O25b:H4 and O1:H7 (Gordon *et al*., 2017; Weissman *et al.,* 2006), and numerous virulence genes such as, *sitA, vat, neuC* and *usp* (Gordon *et al*. (2017).

In humans, this lineage has been implicated in many significant extra-intestinal infections: a study conducted in San Francisco, US, examined *E. coli* isolated from peritoneal fluid and blood samples between 1997 and 2006, and reported that ST95 strains were the most frequently encountered (Bert *et al*., 2010). In addition, ST95 isolates obtained from extra-intestinal infections in clinical and community settings in United States were the second most common, after ST131 (Salipante *et al*., 2015; Bannerjee *et al*., 2013; Adams-Sapper *et al*., 2013). There are also previous reports of interspecies transmission of ST95 strains; it has been isolated from domesticated animals, such as dogs and cats (Johnson *et al.,* 2008), and has been linked with

colibacillosis in birds (Mora *et al*., 2009). However, recent studies rarely observe ST95 strains in companion animals and wild birds *(*Gordon *et al.,* 2017*)*.

Although the ST95 lineage is pandemic, its strains are largely antibiotic susceptible. Unlike what has been continuously observed in the ST131 lineage, only a small proportion of ST95 isolates have acquired multi-drug resistance. In addition, ST95 lineage do not harbour any of the chromosomal mutations in *gyrA* and/or *parC,* which are associated with resistance to fluoroquinolones (Gordon *et al*. 2017).

1.2.3 *Escherichia coli* ST73

ST73 lineage is also spreading globally through clonal expansion, but it is yet to have as much scientific scrutiny as the globally circulating ST131 lineage (Petty *et al*., 2014; Price, *et al*., 2013).

Most of the strains of ST73 have serotype O6:H1 (Martinez-Medina *et al*., 2009; Johnson *et al*., 2008). Other serotypes have also been reported: O2:H1, O18:H1, 18ac:H1, O21:H1, O22:H1, 25:H1, O120:H1 and O120:H31 (Kallonen *et al*., 2017). Numerous virulence genes are also present in this lineage. Such virulence genes include: *cnf1, fimH, fyuA, hra, hlyA, iroN, kpsM II, ompT*, *papC, pic, papEF, papG, sfa/foc, usp* and *vat* (Bert *et al*., 2010; Martinez-Medina *et al*., 2009).

The ST73 lineage has also been implicated in many extra-intestinal sites: in a study conducted between 2005 and 2006, in Rio de Janeiro, Brazil, it was reported that ST73 was the most commonly encountered lineage isolated from urinary tract infections in

women, after ST131 (Dias *et al*., 2009). This report further revealed that the infections were acquired from community sources, affecting about 12% of women. Other studies also isolated ST73 from cystitis and blood stream infections from various clinics and community acquired infections, and at varying frequencies: 5%, 16.6% and 18% (Horner *et al*., 2014; Gibreel *et al*., 2012; Brisse *et al*., 2012).

Although ST73 strains are also known to be pandemic, and are frequently encountered in extra-intestinal sites, they have also been noted to be highly susceptible to antibiotics. It has been reported that ST73 strains are sensitive to most medically important antimicrobials, and do not frequently carry plasmids, which is the typical facilitator of multiple-drug resistance (Bengtsson *et al*., 2012).

1.3 Mobile genetic elements (MGEs).

MGEs enable comprehensive insights into genome evolution. They are genetic elements that can be transferred from one cell to another or move around within a genome. Barbara McClintock first discovered a type of MGE in the 1950s known as transposons. Transposons have the capacity to change location within the genome, cause unstable mutations and also produce chromosomal abnormalities (McClintock, 1950). In the 1970s, another MGE, "insertion sequences" (IS), was identified (Malamy *et al*., 1972).

MGEs are very important elements in horizontal gene transfer (Lukjancenko *et al*., 2010; Touchon *et al*., 2009). They enable AR genes and virulence factors to be transferred within bacterial cell and / or between cells, different lineages, or species

(Partridge, 2011). They are known to enhance the variability and size of genomes across various strains, and also increase host adaptation and fitness to certain environments (Touchon *et al*., 2020; Leimbach *et al*., 2013). Many of these reports have demonstrated that MGEs play crucial roles in antibiotic resistance in pathogens, as well as in the evolution of pathogenic organisms (Stokes and Gillings, 2011; Dobrindt *et al*., 2010; Rasko *et al*., 2008).

1.3.1 Transposable elements

Transposable elements are frequently encountered and are very diverse. Three different types of bacterial transposable elements are found. They include insertion sequence (IS) elements, transposons, and conjugative transposons. These elements, while moving from one location to another can replicate and insert copies of DNA segments within the same cell. *E. coli* strains have been noted to carry multiple copies of transposable elements, IS1 through to IS4. Majority of these MGEs are found on intrinsic plasmids (Muñoz-López and García-Pérez, 2010).

Insertion sequences (IS) are short DNA elements that encode only the genes involved in their own transposition. They are the simplest type of the transposable elements. They are flanked by inverted repeat sequences, which are necessary for the transposition process (Bennet, 1995). They are not directly involved in dissemination of antibiotic resistance because they lack extrinsic information. However, they act as recombination hotspots due to homogeneity of their repeat sequences (Prentice *et al*, 2001). Transposons closely resemble IS elements, except that they encode additional information that are not relevant to their transposition. Transposons consist of a length

of DNA that is flanked on both sides by IS elements, in this situation, the IS elements have captured each other. They then move as a unit with the DNA sequence they flank. The physical closeness of the flanking IS elements is very important to their ability to move as a single unit (Chandler *et al*., 1982). This is because, as a single unit, if an additional DNA sequence is inserted between them, and a threshold size is exceeded, the IS elements would cease to function as a unit. Instead, they will move independently, leaving the earlier associated genes behind. Conjugative transposons carry genes that mediate their integration and excision into and from the chromosome. They also carry genes necessary for their conjugal transfers into bacterial cells (Christopher *et al*., 2014; Salyers *et al*., 1995).

1.3.2 Plasmids

Plasmids are small double stranded DNA molecule found in bacteria and archaea. They are important mobile genetic element that exist independently of the chromosome and are self-replicating. In *E. coli*, plasmids are typically found to comprise less than 5% of the total genome (Martinez-freijo *et al*., 1998). They are known to encode genes required by the cell, especially during adverse conditions. Virulence genes, antibiotic resistance genes, genes resistant to toxic heavy metals and many others enable the cell to exploit specific environmental situations (Thomas and Summers, 2008). Plasmids play a very important role in horizontal gene transfer and most *E. coli* carry two or more plasmids (Sherley *et al*., 2003).

1.3.3 Integrons

Integrons are genetic elements that capture, express and transfer AR genes through their encoded gene cassette *(*Mazel, 2006*).* This process occurs during bacterial conjugation. Integrons are often found in plasmids, bacterial chromosome as well as in transposons (Mazel, 2006). They play a major role in the evolution and adaptation of pathogenic microorganisms.

Based on sequence conservation, integrons consist of 3'-conserved segment (3'-CS), 5'-conserved segment (5'-CS), and the variable segment between 3'-CS and 5'-CS. Integrons are divided into components. The first component is an integron integrase (*intI*) gene. An integron can encode multiple gene cassettes arranged as an array within a conserved recombination site (*attI*), the second component that make up the integron. Integron encoded gene cassettes can move in and out of the array, changing the order of their arrangement within it. These recombination and excision processes are mediated by the *intI* gene (Kovalevskaya, 2002). The third component is an integron connected to a promoter (*Pc*), that is used to express the gene cassettes (Mazel, 2006). Gene cassettes are flanked by conserved sequences known as 59-base element (also known as attC) with one to two open reading frames. As soon as *intI* is expressed, it explores the genomic sequence associated with gene cassettes. It then moves out of the gene cassette and merges into the integron, at the *attI* site. The *attI* allows the separation of any open reading frames (*orfs*) in gene cassettes, with the aid of polymerase chain (PCR) primers, using the attC. This helps to connect novel genes (Boucher *et al*., 2007).

17% of bacterial genomes in the National Centre for Biotechnology Information (NCBI) database harbour an *intI* gene (Cambray *et al*., 2010). The function of integrons in the spread of AR is very significant, especially in Gram-negative bacterial cells. Integrons have acquired an extensive array of resistance genes from reservoir of drug resistant organisms. The rapid increase in integrons that are encoded in AR pathogens may further raise the possibility of interactions with other genes, leading to the acquisition of novel, and complex mobile elements that encode resistance to multiple antibiotics. One of the effects of this is the capacity of MGEs to enhance the transfer of resistance genes within species and cells (Partridge, 2011). Evolution and adaptation of integrons also continues to increase because of persistent exposure of pathogens to selective agents, both in the natural environments and through human activity.

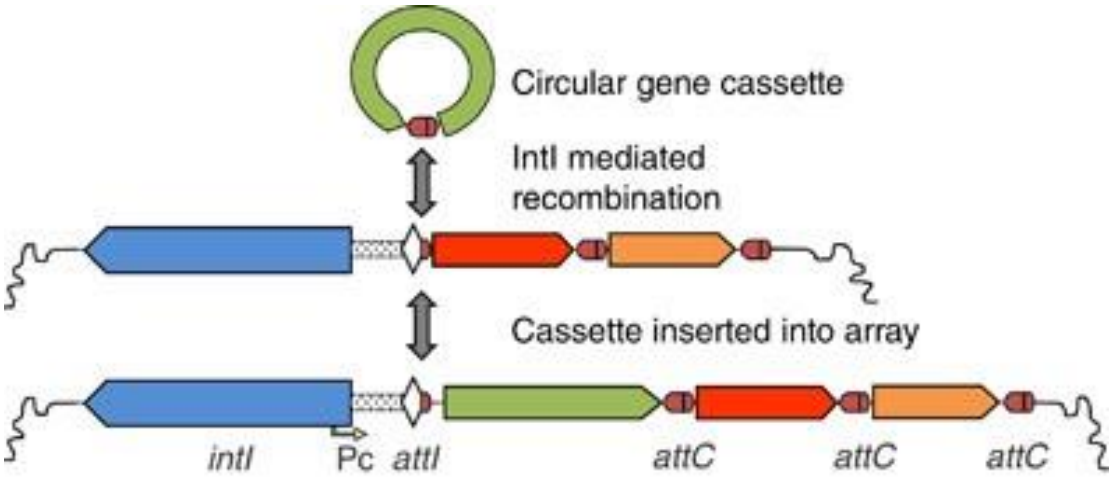

Figure 1.1 Integron structure

Doi: 10.1038/ismej.2014.226

1.3.3.1 Gene cassettes

Gene cassettes are small genetic elements that are composed of a sole or occasionally two genes. They often have a recombination site but lack promoter (Hall and Collis,

1995). Free form of gene cassettes is circular in shape, but when they integrate into integron, they become linear (Hall and Collis, 1995). Integrons may gain one or more gene cassettes to establish an assembly of cassettes (Gillings, 2014). Cassette arrays in clinical class 1 integrons usually consist of between 0 and 6 cassettes (Jacquier *et al*., 2009), while some chromosomal cassette arrays harbour hundreds of gene cassettes (Michael and Labbate, 2010). Gene cassettes encoded within integrons are predominantly key constituents of bacterial adaptation. The capacity of integrons to obtain novel gene cassettes and to reposition the genes already in the gene assembly, enable a faster way to acquire adaptive diversity. The vast diverse functions of gene cassettes harboured by integrons allow the latter to acquire pre-adapted traits for antimicrobial resistance (AMR) gene acquisition and expression, which are very useful in the presence of antibiotics. Therefore, gene cassettes contribute immensely to the adaptation of bacterial cells to antibiotic therapy (Gillings, 2014). Variations in heterogeneous sites as well as in the codon usage in these integrons are proof that accumulation might have been progressive from a variety of phylogenetic environments (Partridge *et al*., 2009).

1.3.3.2 Classification of Integrons

Integrons were initially grouped into mobile integrons and chromosomal integrons, but the two groups were classified based on limited examples. Reports later indicated that there are many other different integron families (Boucher *et al*., 2007; Hall and Stokes, 2004). These integrons are classified based on the phylogeny of their respective integrase genes: the first group is found in proteobacteria from freshwater and soil environments, included in this group are the class 1 and 3 integrons from clinical

samples (Naas *et al*., 2001). The second group is found in gamma-proteobacteria from marine environments, this group includes the class 2 integrons and integrons found on the SXT integrative conjugative element and pRSV1 plasmid from various *Vibrio* species. The last group of integrons are those whose integrase genes are in the reverse orientation to the first two groups (Boucher *et al*., 2007). These integron classes have a diverse reservoir of gene cassettes that are known to harbour resistance to antimicrobials. AR integrons are known to share common characteristics: they are often associated with plasmids, and they have few gene cassettes encoded with AR. But these shared characteristics are not inherent features of their ancestors. It is possible that they might have evolved because of convergent evolution. This type of evolution is often enhanced by strong selection pressures, caused by use of antibiotics by humans. Consequently, all the groups of integrons associated with AR pathogens may have current and same evolutionary record (Gillings, 2014). An example of this, is the *intI* genes associated with the integrons from the class 1 mobile integron. These *intI* genes were isolated from various hospitals and were observed to be similar. This suggests that these integrons are new progenies of a single event (Gillings, 2014).

However, both chromosomal integrons and mobile integrons play crucial function in genome evolution and diversity (Boucher *et al*., 2011). Mobile integrons have limited gene cassettes, with the longest registered being about 8 cassettes driven from a single promoter (Naas *et al*., 2001). Additional cassettes are rarely expressed because of their distance from the promoter (Hall and Collis, 1995). Reservoir of cassettes harboured by mobile integrons, can confer resistance to most categories of antimicrobials used in agriculture and medical science (Partridge *et al*., 2009; Gillings

et al., 2008; Stokes *et al*., 2006). They also have varying *attI* and are able to gain their mobility by interacting with plasmids or transposons. The second group known as the chromosomal integron is also referred to as “super-integrons”. This group has hundreds of gene cassettes, and similar *attI* that are often not involved in AMR (Gillings *et al*., 2008). They are said to be inactive (Cambray *et al*., 2010), although movement of gene cassettes from chromosomal integrons has been reported (Domingues *et al*., 2012). Generally, integrons do not have capacity for self-mobility; the *intI* is often unable to remove its own gene from a chromosome. Such genes are usually located within cassette arrays of chromosomal integrons or are found lying adjacent to the *intI* gene. The *intI* thus, have to depend on their association with transposases or recombinases for inter-chromosomal mobility.

1.3.3.3 Class one Integrons (C1)

C1 integrons harbour limited gene cassettes, and they usually participate actively in the spread of AR. There have been numerous reports of plasmids harbouring C1 integrons, in a broad array of bacterial species (McGivern *et al*., 2021; Li *et al*., 2020; Wang *et al*., 2017; Van Essen-Zandbergen *et al*., 2009).C1 integrons are found in about 40 to 70% of Gram-negative pathogens from clinical samples (van Essen-Zandbergen *et al*, 2007; Martinez-Freijo *et al*.,1998). The movement of Tn402-integron hybrid into diverse plasmids and other transposons generated diversity and rapid spread of C1 integrons into a range of pathogens and commensals. The C1 clinical integrons are very commonly encountered. They have been discovered at about 10 to 50% in commensal bacteria in healthy human subjects (Liu *et al*., 2013; Bailey *et al*., 2010), where integron carriage by commensal *E. coli* increased to about 80% (Sepp

et al., 2009).Commensal bacteria encoding integrons act as a bridge for horizontal gene transfer of resistance genes between the commensals and pathogens (Djordjevic *et al*., 2013; Betteridge *et al*., 2011). Although this class of integrons are commonly encountered in Gram-negative bacteria, some studies have also detected their presence in Gram-positive bacterial cells (Xu *et al*., 2011; Gillings *et al*., 2008; Shi *et al*., 2006; Nandi *et al*., 2004). C1 integrons that are enclosed in transposons or other mobile genetic elements, have been shown to have the capacity for external mobility between and within bacterial genomes. This process is known as hitchhiking. C1 integrons encoded in plasmids are often transferred during bacterial conjugal gene transfer, or through natural transformation or transduction (Domingues *et al*., 2012).

1.4 Horizontal Gene Transfer and Mobile Genetic Elements

The transfer of genetic material amongst prokaryotes and/or eukaryotes is called horizontal gene transfer (HGT) (Robinson *et al*., 2013; Hotopp, 2011; Ochman *et al*., 2000). This process plays an important function in the increase of bacterial resistance to antimicrobials (Gyles and Boerlin, 2014; Barlow, 2009). This occurs when a bacterial cell gains resistance, and such resistance genes are spread to other cells (Francino, 2012; Hawkey and Jones, 2009). Hence, horizontal gene transfer is important in evolution and adaptation of organisms (Gyles and Boerlin, 2014; Francino, 2012; Organisation for Economic Co-operation and Development (OECD), 2010; Hawkey and Jones, 2009; Kay *et al*., 2002; Koonin *et al*., 2001; Nielsen, 1998). The mechanisms of horizontal gene transfer, powerful natural selective pressures and transposition have resulted in multi-drug resistance species of most bacterial pathogens (Gyles and Boerlin, 2014; Stearns and Hoekstra, 2005).

Approximately one percent (1%) of genes that have been horizontally transferred are commonly found in areas known as hotspots within the chromosome. Increase of these genes could occur with increased genome size and rate of plasmid transfer, thereby enhancing the enlargement of the hotspots. Hotspots direct most changes on gene banks. They are home to most MGEs, especially those encoding AR genes, and also control exchange between genome spread and assembly (Oliveira *et al*., 2017).

1.4.1 Mechanisms of Horizontal Gene Transfer

AMR genes can be transferred from one bacterium to another through three different methods of horizontal gene transfer. They are transformation, transduction and conjugation (Gyles and Boerlin, 2014; Todar, 2012; Maloy, 2002).

1.4.1.1 Transformation

Transformation involves the introduction, absorption, and expression of genetic material (usually DNA or RNA) from the environment, into a competent recipient cell, leading to the genetic modification of the cell (Stearns and Hoekstra, 2005). Competency is a transient situation, whereby a recipient cell has the capacity to accommodate foreign DNA from the surroundings or be induced in a laboratory setting (Johnston *et al*., 2014). This state is a ‘must’ in order to integrate the genes into bacteria for experimental studies or for clinical or commercial purposes. Bacteria are usually the commonest organisms of interest for the transformation process (Renner and Bellot, 2012). Transformation can also take place naturally as a time-limited reaction to adverse situations, such as, cell density and starvation (Johnston *et al*., 2014). In this case, plasmid DNA harbouring a beneficial determinant gene, moves

from one bacterial cell to another, where it is joined to the genome of the recipient cell by homologous recombination.

1.4.1.2 Transduction

Transduction occurs when foreign DNA is introduced into a cell by a virus (Snyder *et al.,* 2013*;* Stearns and Hoekstra, 2005*)*. Bacteriophages (phages) are viral elements that invade bacterial cells and reproduce within the cells. They harbour resistance genes, especially β-lactamase genes, which are MDR genes, as well as virulence factors needed by the bacteria. This unique role usually results in bacterial evolution and adaptation. Molecular biologists utilise the transduction process to transfer foreign DNA into cell genome of a host, with the aid of a bacteriophage. Transduction involves any of two cycles: the lytic cycle or lysogenic cycle. Lytic cycle occurs when bacteriophages infect bacterial cells. They utilise transcriptional, replicational and translational machineries of the host cell to produce new viral particles known as virions. The virions lyse the host bacterial cell, and then get discharged. However, in the lysogenic cycle, after infection, the chromosome of the phage gets embedded as a prophage (phage genome) into the chromosome of the bacterial cell, where it could become passive for a long period of time. It can only become active again, in a situation where it is stimulated, for example, by ultra-violent light. If this occurs, the prophage excises from the chromosome of the bacterial cell, and then attacks the cell, for the lytic cycle to take place (Snyder *et al*, 2013; Malys, 2012; Madigan and Martinko, 2006).

1.4.1.3 Conjugation

Conjugation is the movement of genetic material among bacterial cells, either by direct contact between one cell and another, or by a bridge-like connection between the two cells (Johnston *et al*., 2014; Stearns and Hoekstra, 2005), with the aid of a pilus (Alberts *et al*., 2002). Although this process may be via direct contact between one cell and another, or by means of a pilus, there is no exchange of gamete. Therefore, a new organism is not formed. Nevertheless, an extant organism is transformed. In the course of a traditional *E. coli* conjugation process, a conjugative or mobilisable genetic material, such as a plasmid or transposon is usually supplied by the donor cell (Keeling and Palmer, 2008). Such genetic information is very often useful to the recipient cell, and some of the advantages are antimicrobial resistance, capacity to utilise novel metabolites, virulence, and ability to tolerate xenobiotic chemicals (Gyles and Boerlin, 2014).

1.5 Plasmid diversity

1.5.1 Plasmid classification

1.5.1.1 F-plasmids

The F-plasmids also known as fertility plasmids are one of the earliest plasmids responsible for antibiotic resistance. They are frequently found in the family Enterobacteriaceae. These plasmids allow genes to be transferred from one bacterium, carrying the F-factor, known as donor, to another bacterium that is devoid of the F-factor (F^-) known as recipient, by conjugal transfer *(*Lawley *et al., 2003)*.

The F-factor is a 100kb fragment of double stranded DNA with two origins of replications (Rosenberg and Hastings, 2001). Transfer origin, *oriT*, enables a specific one-way direction replication during horizontal gene transfer. The F-factor harbours genes for sexual pilus. The pilus forms a bridge-like connection, used for F-carrying bacteria, to attach to the F^- cells for conjugative transfer. The F-factor also encodes an operon of about thirty genes, which are known to encode *tra* proteins, that enhance gene transfer. The mating apparatus of F-type plasmids are also found to have different replicons. Such subgroups of F-plasmids, including FIA, FIIA / FIC and FIB are commonly encountered (Mutai, *et al*., 2019; Rosenberg and Hastings, 2001). These replicons carry fertility plasmids necessary for bacterial conjugation. In addition to multiple replicons, most F-plasmids also carry varying toxin-anti toxin systems as well as partitioning systems. These later systems ensure plasmid replication and maintenance (Van Melderen and Saavedra, 2009; Lewis-Kittell and Helinski, 1993).

1.5.1.2 Resistance plasmids (R)

R plasmids are plasmids known to encode AR genes, often multiple AR genes (Levin, 1995). The multi-resistance of R plasmids can be described by their capacity to spread numerous AR genes at once through bacterial conjugation. Most F-plasmids are transferred at a very high frequency, unlike R plasmids that are usually transferred at low frequencies. In addition, some R plasmids when they co-exist in the same host cell with F-plasmids, prevent the transfer of F-plasmids (Watanabe, 1969).

1.5.2 Plasmid Incompatibility

Bacterial cells may harbour multiple copies of a plasmid. But not all plasmids are capable of stable co-existence within the same cell. Two or more co-resident plasmids are not able to co-exist when they both share same origin of replication or similar partitioning system. This situation is termed plasmid incompatibility (Carattoli, 2009). In plasmid incompatibility, the plasmids contribute equally to the total plasmid copy number and are thus regulated together. However, they are not identified as distinct plasmids. Consequently, during cell division, one of the plasmids usually gets lost (Novick, 1987). This situation is common with low copy number plasmids that belong to the same incompatibility (Inc) group.

Plasmid incompatibility can either be reciprocal or unidirectional. Reciprocal incompatibility is when the two plasmids have an equal likelihood of getting lost from the bacterial cell they began with, while unidirectional incompatibility occurs when one of the plasmids has an extra attribute that gives it an edge over the other one, namely, acquisition of a second replicon (Thomas, 2014).

Well-established classification systems have been used to group incompatible plasmids. Unlike previous replicon schemes that are based on observational analysis (Datta and Hedges, 1973), now, incompatibility classification is based on amino acid sequence of replication initiation (Rep) protein (Shintani *et al*., 2015). This plasmid classification type utilises molecular characteristics of the replicons (Carattoli *et al*., 2005). However, there are disadvantages to this classification type: it is difficult to group multiple replicon plasmids into single replicon groups, there is minimal information about the Rep types and the Inc groups of many microbial taxonomies and

also, the difficulty in distinguishing the replication sites for other types of plasmids. Classification of incompatible plasmids based on mating pair formation (MPF) classes and mobility (MOB) types have been reported (Garcillan-Barcia *et al*., 2011; Smillie *et al*., 2010) to solve these challenges.

Twenty seven (27) Inc group plasmids presently exist in the Enterobacteriaceae (Shintani *et al*., 2015). Plasmids that exhibit extensive dissimilarity from the same progenitor display compatibility and therefore can co-exist with one another while plasmids that are nearly similar tend to be incompatible. Based on this genetic relationship, Inc group plasmids were further classified into four key groups: IncI group (IncB, IncK, IncI), IncF group (IncC, IncD, IncF, IncJ, IncS); IncP group (IncM, IncP, IncU, IncW) and Ti group (IncH, IncN, IncT, IncX) (Waters, 1999). The Inc group plasmids F, N, P, B/O, HI1, HI2, I1, IncA/C and L/M are all associated with clinically acquired MDR infections. The IncP-1 plasmid is a well-studied plasmid that is distributed across many bacterial classes (Adamczyk and Jagura-Burdzy, 2003). Rep proteins of IncB, IncF, IncI, and IncK plasmids are more than 87% identical to one another. IncF plasmids especially IncFII and IncFIB are the most prevalent in *E. coli* (Wang *et al*., 2013; Carattoli, 2009).

1.5.3 Plasmid Copy Number

Plasmid copy number is the number of plasmid copies in a bacterial cell. In most plasmids, there are one or two copies per chromosome. Low copy plasmids are 5 or less copies in a bacterial cell. Such plasmids often need a toxin-antitoxin or a partitioning system to ensure that each daughter cell receives the plasmid at cell

division. Medium copy plasmids are often mutated plasmids. There are about 20 copies per cell. Mutated plasmids usually replicate to high copy plasmids. Most high copy plasmids are >100 copies per cell and are often derived from medium copy number plasmids through mutagenesis (Boros *et al*., 1984). In some small-size plasmids, like the ColE plasmids, as much as fifty, or more plasmids are often found (Carattoli, 2009). Plasmid copy number plays a great impact in the effectiveness of plasmid-borne attributes, particularly on drug resistance. Most multi-copy plasmids are small, and do not typically encode antibiotic resistance genes (Carattoli, 2009).

Plasmid copy number is intrinsically linked to plasmid incompatibility (Carattoli, 2009). At cell division (fig. 1.2), once replication is initiated, in a cell that harbours two incompatible plasmids, the plasmids cannot be distinguished from one another because they share same origin of replication. Replication is thus delayed until the plasmids have been segregated into different cells, in order to create the pre-replication copy number. Plasmids loss at cell division, as well as the effects of the presence of plasmids on their hosts, can be controlled through copy number regulation (Carattoli, 2009). This process of plasmid copy number regulation are very important and often vary.

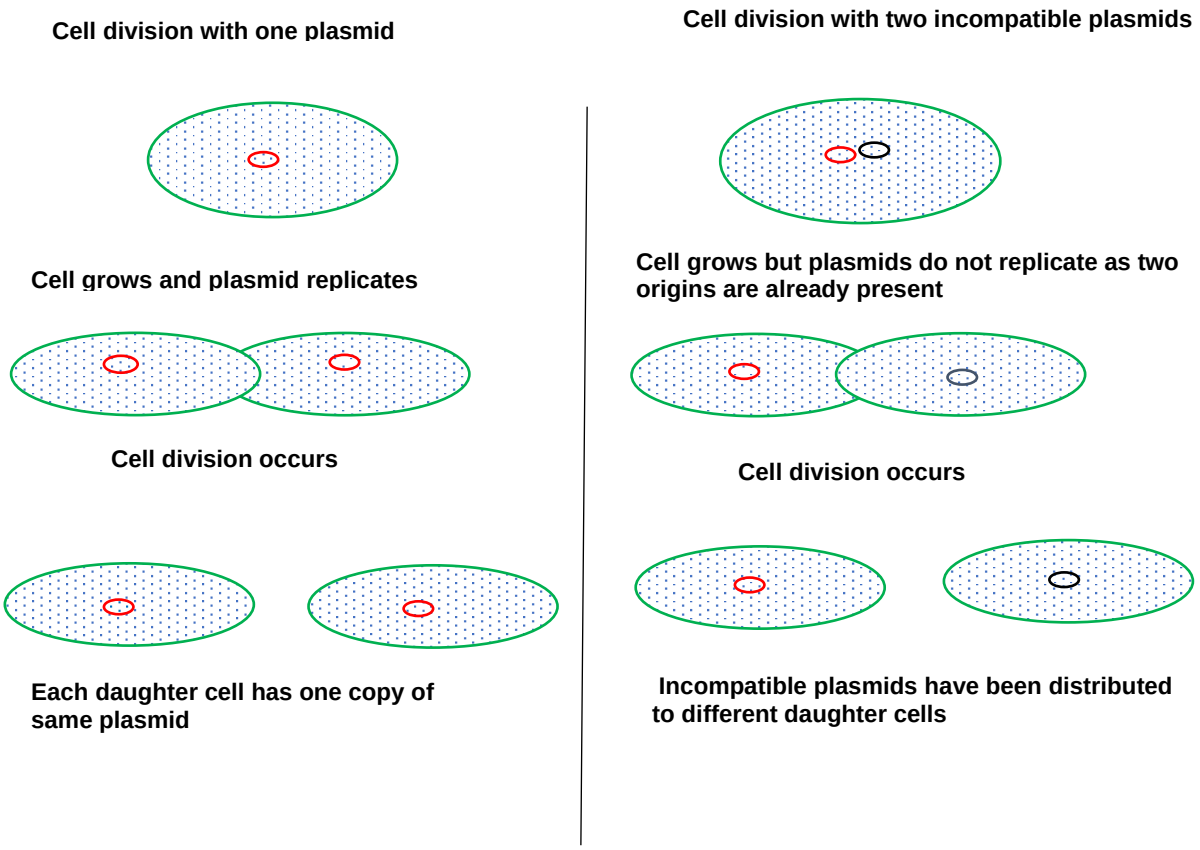

Figure 1.2 Copy number and plasmid incompatibility

Adapted: www.bitesizebio.com

1.6 Plasmid replication control

The number of plasmid copies within a cell is pre-determined for a given plasmid under constant conditions (Lewis-Kittell and Helinski, 1993). This condition is accomplished by controlling plasmid replication, through control systems, as well as in the regulation of the replication-initiation gene harboured by the plasmids. Most of plasmid control mechanisms involves restriction (Lewis-Kittell and Helinski, 1993). Restriction refers to a negative feedback control that ensures an inverse relationship in the rates of replication and plasmid copy number in each cell. The region of plasmid that accomplishes the initiation of replication, and its control is called a replicon. The replicon consists of replication origin and gene-identifying proteins that are frequently

needed for the integration and replication sequences, required in the control of replication. During bacterial cell division, replicon encode genes that regulate uniform partitioning of plasmids into daughter cells (Lewis-Kittell and Helinski, 1993).

Replication control can be achieved either by ribonucleic acid (RNA) antisense transcripts or by iteron based control. RNA antisense transcripts controls plasmid replication by an inhibition process: RNA is transcribed in a region overlapping an RNA transcript that is either directly involved in the replication process or encodes a product that is involved. The two RNA molecules hybridise along their complementary sequence regions, preventing further action from RNA transcript normally involved in replication. Counterscript RNA molecules are constitutively produced, hence their concentration within the cell, and the probability of interfering with plasmid replication, which is proportional to plasmid copy number (Brantl, 2014; Lewis-Kittell and Helinski,1993). An example of this is RNA antisense transcripts detected in IncFII plasmids. In iteron based control, a diffusable initiator protein attaches to sets of tandem repeats sequences known as iterons at the origin of replication. The association between the iteron and replication initiator protein enhances plasmid replication and control. At low concentrations, binding of the protein initiates the process of replication. But at high concentration, initiation protein will become inhibitory as excess binding obscures the origin of replication (Chattora, 2000; Lewis-Kittell and Helinski, 1993). An example of iteron-based control is observed in IncP plasmids (Stalker *et al*., 1981).

1.7 Plasmid Maintenance Systems

1.7.1 Post-segregational killing (PSK) systems

Post-segregational killing (PSK) systems (fig.1.3), also known as toxin / antitoxin (TA) systems or plasmid addiction systems, consist of small closely related genetic elements, that encode a stable "toxin" protein and a sensitive "antitoxin". PSK systems are diverse and are often prevalent in multiple copies in the prokaryotes (Fozo *et al*., 2010). The activities of this system have been expressed as programmed cell death. At cell division, PSK is hindered by the presence of a plasmid in a cell, usually because the antitoxin immobilises the toxin, to prevent the killing. However, if there is plasmid loss, often because of replication errors, or mis-segregation (Yarmolinsky, 1995), the antitoxin quickly breaks down, and the toxin kills the daughter cells that do not maintain their plasmid copies (Hayes, 2003; Gerdes *et al*., 1986), that is, the plasmid-free bacteria are eventually eliminated by the adverse effects of the toxin (Wang and Wood, 2011; Engelberg-Kulka *et al*., 2006; Tsuchimoto *et al*., 1992). Another system of cell killing occurs when the antitoxin acts as an antisense transcript and blocks the function of the plasmid messenger RNA (mRNA) to act as a toxic protein. In this situation also, if there is plasmid loss, the antitoxin's RNA quickly breaks down leading to the production of the toxin mRNA, which would eventually kill the bacterial cell (Gerdes and Wagner, 2007; Gerdes *et al*., 1997). PSK systems usually move from one cell to another through the process of conjugal transfer, and they sustain themselves at the detriment of their host cell (Van Melderen and Saavedra, 2009).

The PSK system is associated with pathogenic bacteria, and as such, can harbour plasmids that encode AMR and virulence factors (Ma *et al*., 2013). *Hok* and *sok* are a type 1 PSK system detected in the plasmids of a number of Gram-negative bacteria,

including plasmid R1 of *E. coli* (Fozo *et al*., 2008a)*. FimA* and *FimB* are homologues of *hok* and *sok.* They stabilise the F-plasmid in *E. coli*. The *hok* toxin is found in the cell membrane, and often functions by killing the membrane capacity of the bacterial cell, resulting in the depletion of the cell's fitness needed for metabolism. The antitoxin *sok*, on the other hand, is a small volatile antidote that blocks the toxin *hok,* by killing its movement (Gerdes *et al*., 1986). In addition, other PSK systems, *ccdB* and *ccdA* are also found in the F-plasmid of *E. coli* (Fozo *et al*., 2008b) while *Ibs* and *Sib* are found in the intergenic regions of *E. coli* (Bernard and Couturier, 1992). PSK systems were also detected on other transposable elements, such as bacteriophages and conjugative transposons (Magnuson, 2007*).* They have also been detected in multiple copies in chromosomes (Alonso, 2021; Harms *et al*., 2018; Brantl, 2012; Van Melderen, 2010).

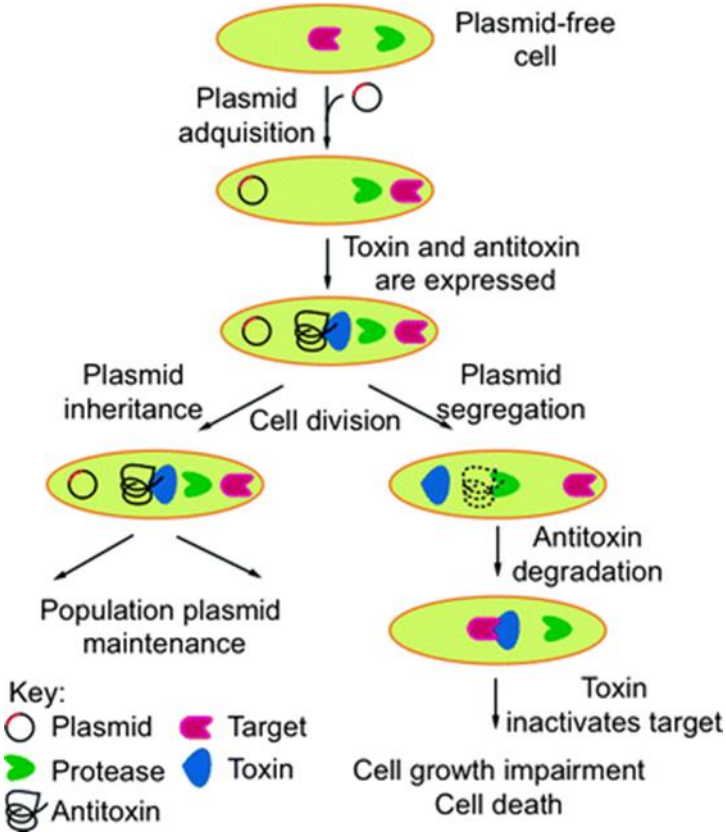

Figure 1.3 Post-segregational killing systems

Source: https://images.app.goo.gl/AZFQjSgVqhMBkMp97

1.7.2 Multimer Resolution Systems

Homologous recombination and replication between sister plasmids, can lead to plasmid dimers (two plasmids, each with its own replication origin) or plasmid multimers (sequential repeats of monomer plasmids). Either of these could lead to a reduction in the number of plasmid copies that are available for segregation into daughter cells (Austin *et al*., 1981). In large plasmids (often available as two or three copies for every bacterial cell), plasmid dimers or multimers could lead to recurrent plasmid loss (Austin *et al*., 1981; Nordström and Austin, 1989). Furthermore, the type of replication control that is utilised by most plasmids, may lead to what is known as “dimer catastrophe”. This situation occurs because dimers or multimers contain cells that are slow growing and are also maintained at low copy number than monomers (Summers *et al*., 1993). The likelihood of monomer replication is thus lower than that of dimer or multimer, resulting in increased proportion of dimer only cells. The effect of this increase is the segregation of cells that are devoid of plasmid. Dimer catastrophe often leads to increased reduction in plasmid stability (Summers *et al*., 1993).

Multimer resolution systems help in improving the number of plasmid copies available for segregation, by correcting plasmid dimers and multimers (Austin *et al*., 1981). Before cell division, it is very necessary that multimers are rectified to monomers for effective segregation to occur. This is achieved by site-specific, enzyme-mediated, recombination systems. Large plasmids harbour their own recombinase systems, which is composed of genes for a recombination site and a distinct recombinase. Dimers are composed of two of such sites. These sites are nicked, changed, and re-assembled by the protein to produce two discrete circular monomers.

1.7.3 Partition (*par*) systems

Large plasmids are composed of low copy numbers, usually two or more copies in a cell. Irregular segregation of these plasmid copies into daughter cells can result in unrestricted plasmid loss (Nordström and Austin, 1989). In this situation, PSK can eliminate a great number of the cells within the population. This is a disadvantage to the survival of the pathogenic cell. Plasmids can overcome this situation by means of plasmid *par* systems.

Most conjugative plasmids with low copy numbers harbour a *par* system. This system consists of a *cis*-acting centromere-like region, an ATPase or a GTPase motor protein and a distinct DNA binding adaptor protein. The transcription of the operon harbouring the two proteins is often auto-regulated (Gerdes *et al*., 2000). Plasmid *par* systems can be grouped into three classes: Type I, II and III. Type I *par* system harbours a divergent Walker-type P-loop ATPase. This class of *par* systems can be additionally grouped into *Ia* and *Ib* (Gerdes *et al*., 2000). Type *Ia* system have large ATPase motor proteins, which functions as the operon self-repressor. Type *Ib* systems have small motor proteins and use their DNA binding protein as a self-repressor. Type II *par* system harbours an actin / hsp70 class of ATPase, while type III *par* system harbours a GTPase (Gerdes *et al*., 2000).

Par systems separate sister plasmids into daughter cells by a method comparable to the mitotic chromosome segregation (Ebersbach and Gerdes, 2005). The segregation of low copy plasmids at cell division generally depends on these systems. Plasmid *par* system ensures that one copy of the plasmid DNA is always acquired by each daughter cell (Gordon and Wright, 2000). In a situation where *par* system is absent, successful

plasmid segregation will depend on, if the plasmid is a multi-copy plasmid, and the actual spread of the plasmids within the cell (Wang *et al*., 2016; Reyes-Lamothe, *et al*., 2014).

1.8 Population Biology of Plasmids

1.8.1 Conjugative systems

Plasmids are transferred among lineages through conjugation and mobilisation process (fig.1.4). At the end of this process, each of donor and recipient cell are left with a duplicate of the plasmid (Griffiths *et al*., 2000). Plasmids encoded by conjugative transfer systems harbour specific genes required to facilitate DNA exchange (Cabezón*, et al*., 2015; Derbyshire and Gray, 2014; Gomis-Rüth and Coll, 2006; Schröder and Lanka, 2005; Lawley, *et al*., 2003; Llosa, *et al*., 2002; Haase, *et al*., 1995). Conjugative transfer systems make sure that the same genetic element is not present in the recipient cell. This is achieved by blocking the formation of cell-to-cell contact, and the entry of single stranded DNA into the recipient cell. Thus, preventing the transfer of plasmid DNA into cells already with similar conjugative plasmid.

The conjugation system involves three major constituents: a relaxosome, a type IV secretion system (T_4SSs) and a type IV coupling protein (T_4CP). The relaxase, is a protein found within the relaxosome. It identifies the origin of transfer (oriT), which is a small DNA sequence used only for bacterial conjugation (Smillie *et al*., 2010). The relaxase catalyses the initial and the final stages in the bacterial conjugation by increasing the initial break at the origin of transfer within the donor cell. This process occurs when the relaxase cuts the double stranded DNA and hitches the emerging

single stranded DNA at the transfer origin. The relaxase, now attached to the single stranded DNA molecule, is incorporated by a T_4CP into a T_4SSs, and then it moves from the donor cell to the recipient cell (de la Cruz *et al*.**,** 2010).

Most T_4SSs are not associated with bacterial conjugation, but are differentiated in protein secretion, especially in eukaryotic organisms, where the distribution of effector proteins to the cytosol of the organisms occur. T_4SSs do not have a relaxase, but they encode a T_4CP (Smillie *et al*., 2010). T_4SSs comprises eight massive clades that are related to the eight mating pair formation groups. These mating pair formation groups are based on the phylogeny of VirB4, an ubiquitous protein with identifiable homologues (Guglielmini *et al*., 2014). Conjugative plasmids are usually large. Some that are as large as 100 kb or more have been observed (Smillie *et al*., 2010). The size is usually proportional to the amount of plasmid DNA required for the replication process to occur (Smillie *et al*., 2010).

1.8.2 Mobilisable systems

Mobilisable plasmids are small, usually about 10kb. They are known to harbour only a few genes, and also lack the genes that encode the factors needed for the coupling of the cells prior to DNA transfer. However, mobilisable plasmids encode genes that are necessary for their own conjugal DNA transfer (Garcillán-Barcia *et al*., 2009). This conjugal transfer, however, is only possible when a conjugative plasmid co-exist with a mobilisable plasmid within the same cell. Mobilisable plasmids do harbour a relaxase and sometimes, a T_4CP, but they have no T_4SSs. They possess the relaxosome constituents, which is, a relaxase protein, the origin of transfer and some accessory

proteins (Smillie *et al*., 2010). They are found in excess in the genomes of bacterial cells because they utilise a T_4SSs, that is found within the conjugative plasmids that they co-exist with *(*Francia *et al.,* 2004). Certain bacterial cells do not encode either conjugative or mobilisable plasmids. These are known as non-mobilisable plasmids (Smillie *et al*., 2010).

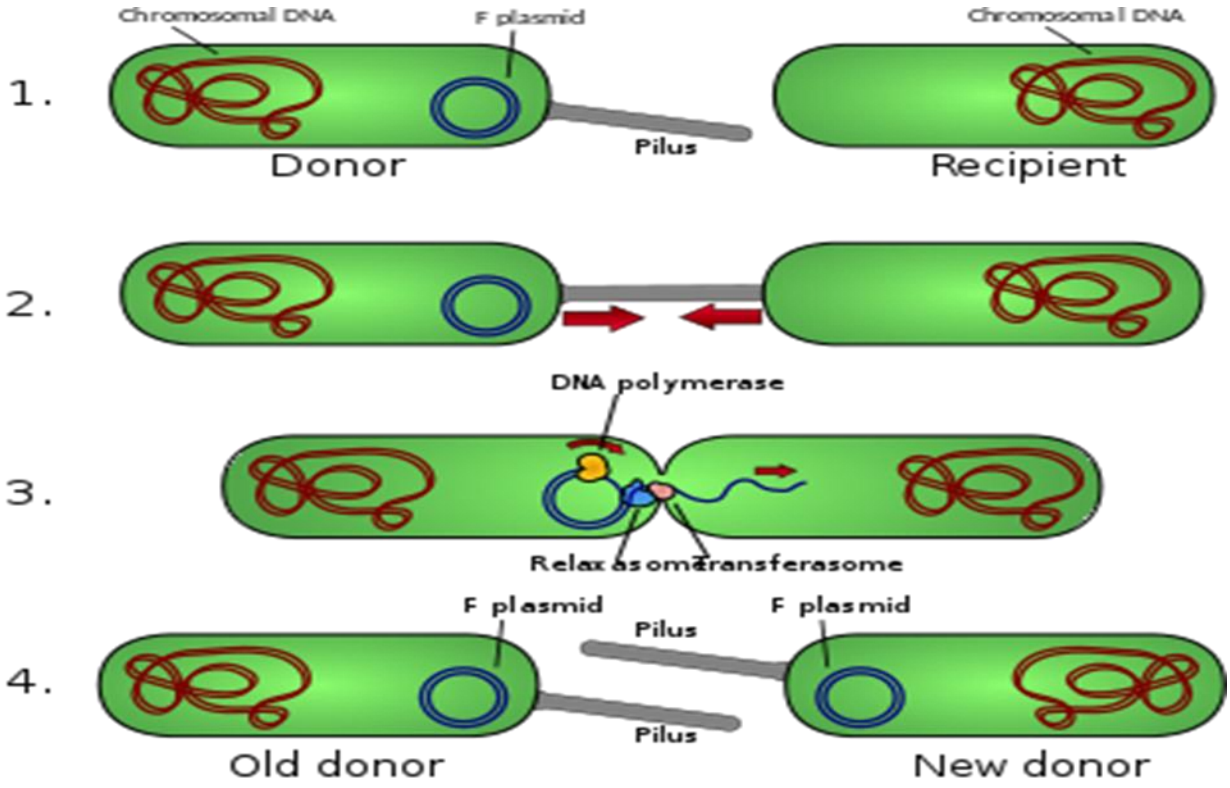

Figure 1.4 Bacterial Conjugation

Source: en.wikipedia.org

1.9 Factors contributing to the persistence of plasmids in the host populations

Plasmids play a very crucial role in the evolution of bacteria, however, understanding how far and how long plasmids can stably persist in the bacterial population remains a difficult task (Lili *et al*., 2007; Bergstrom, *et al*., 2000). This is usually because of the

burden created by the presence of the plasmid(s) within the host cell. This burden can be in the form of plasmid metabolism, certain regulative interchanges between the bacterial cell and the plasmid, cell-killing role acquired from harbouring the plasmid and the rate of plasmid loss during cell division (Baltrus, 2013; Shachra *et al*., 2010; Doyle *et al*., 2007). An equilibrium between the effects of the plasmid and the host fitness is thus, essential for both the plasmid and the bacterial host. There are three main factors that influence the persistence of plasmids in their host populations: plasmid effects on host fitness, plasmid loss due to segregation and fitness gains due to plasmid transfer.

1.9.1 Plasmid effects on host fitness

Host fitness often depend on the costs and the benefits of the plasmids to the bacterial host. Plasmid evolution can be influenced by the number and plasmid DNA content within a host cell. Plasmids exert a load on the bacteria that harbour them. When the plasmid DNA content is large and / or the plasmid has a high copy number, there is a negative effect on the bacterial cell (Smith and Bidochka, 1998; Khosravi *et al*., 1990). This is especially crucial where there is an absence of encoded traits that benefit the bacteria in such plasmids.

Plasmid loss often occurs when the costs of carrying the plasmid exceeds its usefulness and there is negative selection. However, for beneficial plasmid genes (where plasmid trait confers an advantage), integration would be expected (positive effect) thereby minimizing random loss during cell division. Several reports have shown that in some situations, where plasmid-borne determinants confer a benefit to their

hosts, the overall fitness effects of such plasmids are not often well understood (Di Luca *et al*., 2017; Porse *et al*., 2016; Johnson *et al*., 2015; Vogwill and MacLean, 2015; San Millan *et al*., 2014). Some of these studies indicated that the carriage of plasmids encoding antibiotic resistance pose a cost to the bacterial host without the presence of antibiotic. Under these conditions, it is anticipated that bacterial strains that are devoid of the costly plasmid will surpass the plasmid-bearing strains because of the fitness cost of plasmid carriage. Costly plasmids may therefore be lost from the bacterial population because of competition with cells that are plasmid-free. Persistence of such costly plasmids within the cell, would therefore depend on the ratio of the beneficial outcome, to the cost effects on the cell (Steward and Levin, 1977). Other studies have suggested that reduction in the plasmid cost can occur after some time (Starikova *et a*l., 2013; Dionisio *et al*., 2005; Dahlberg and Chao, 2003).

1.9.2 Plasmid loss due to Segregation

Plasmid stability is a term used when each of the daughter cells possess a single plasmid at cell division. Such culture is said to be segregationally stable. This is because all the plasmids have the right base sequence (Friehs, 2004). How well plasmids are maintained determines the plasmid stability. Stability of plasmid is enumerated by calculating the density of cells that are devoid of plasmid over a period, usually, in an already prepared cell culture containing plasmid. The rationale behind this type of experiment is that plasmid losses and any other ensuing positive growth of plasmid devoid cells account for the total number of plasmid-free cells at any period (Lili *et al*., 2007). Even though, at the beginning, this number may be small, but, at every generation, the growth increases. Plasmid persistence can increase, following

adaptive evolution of the plasmid or the bacterial host. However, prolong plasmid persistence within an adapted bacterial cell will depend on the plasmid effect on the host fitness (Wein *et al*., 2019).

Plasmid segregation signifies bacterial plasmid replication, where similar copies of plasmids are generated and are equally segregated among the new daughter cells. Plasmid loss due to segregation is usually because of unequal distribution of the plasmid molecules during cell division. When this occurs, heterogeneous population of cells (where plasmid-free cells outnumber plasmid-bearing cells) emerge. Hence, there is reduction in the overall productivity of the cell population (Popov *et al*., 2009).

Plasmid size and copy number also play important part in segregational loss. High-copy number plasmids are known to be irregularly spread among daughter cells during cell division, because they do not have sufficient partitioning and toxin-antitoxin system (Wang, 2017; San Millan *et al*., 2016; Million-Weaver and Camps, 2014), and because of their high copy numbers, probability of both of the daughter cells having a plasmid at the same time is very low. In addition, when copy number of the plasmid increases, increase in the number of Iterons (repeated sequences within the origin of replication that are needed for replication) often inhibit plasmid replication (Novick, 1987). Low-copy number plasmids on the other hand, need certain systems to be able to disseminate plasmids. Such systems include: replication control mechanism, regulation of copy number and PSK systems (Tsang, 2017; Carattoli, 2013). To establish copy number maintenance, PSK system enhances elimination of cells that are devoid of plasmids and in some cases, also increase plasmid replication (Dmowski and Jagura-Burdzy, 2013; Hayes and Van Melderen, 2011). When plasmid copies are

partitioned irregularly among daughter cells, or when the plasmid replication rate is inadequate to match host replication rate, then cells devoid of plasmids will emerge again in populations of plasmid encoding cells (Lenski and Bouma, 1987; Kline, 1985; Scott, 1984). But, if plasmid carriage inhibits the cell growth, then selection will facilitate the increase in the prevalence of the plasmid-free cells (Warnes and Stephenson, 1986; Seo and Bailey, 1985; Lee and Edlin, 1985; Noack *et al*., 1981). Persistence of plasmids in a bacterial population, thus, will depend on how stable the plasmids are within their host, the fitness of the bacterial host and the environmental conditions of the plasmids (Wein *et al*., 2019).

1.9.3 Fitness gains due to plasmid transfer

Conjugal gene transfer is enhanced both by mobile genetic elements and the bacterial hosts: donor (Frost and Koraimann, 2010) and the recipient cells ((Marraffini and Sontheimer, 2008; Bertani and Weigle, 1953). Most MGEs are advantageous to their hosts via the sharing of the same gene pool. Conjugal transfer could prove advantageous to the host due to the transfer of variable genes that are not directly associated in the maintenance and transfer of MGEs (Dimitriu *et al*., 2016). But the conjugal transfer of these traits has been explained as an exploitative characteristic of the MGEs because of the plasmid cost effects on the host. Although, it has been established that the donor cell does not obtain any direct gain from transferring genes to the recipient cell, however this process is beneficial to the recipient cell (Dimitriu *et al*., 2016). Nevertheless, the plasmid being transferred still enhances the donor allele overall fitness, because the same cell is improved with advantageous allelomorphs (Dimitriu *et al*., 2016). Fitness gains resulting from the transfer of plasmids also select

for resistance and virulence factors, therefore plasmid transfer rates can be enhanced by both indirect plasmid effects on host fitness and direct plasmid effects on plasmid fitness (Stewart and Levin, 1977).

1.10 Persistence of non-mobilisable plasmids in bacterial population

Bacterial conjugal transfers are necessary for plasmid persistence within bacterial population, but reports have shown that half of known plasmids are neither conjugative nor mobilisable (San Millan *et al*., 2013; Smillie *et al*., 2010). Therefore, another question that arises is, how do so many of these non-conjugative and non-mobilisable plasmids continue to persist within the bacterial population. San Millan *et al*. (2014), discovered that when bacteria are exposed to antibiotics, they evolve compensatory adaptations to alleviate some of the effects of carrying a plasmid, and even when antibiotics use is ceased, or when they are exposed to antibiotics again, these adapted plasmid-carriers become positively selected (San Millan *et al*. 2014). Therefore, such a plasmid are much more slowly removed from the population because it is no longer a burden to the bacteria. These findings suggest that adaptations between plasmid and bacteria in the presence of antibiotics, during the first generations of co-existence are enough for stability and maintenance of a non-mobilisable plasmid in a bacterial population, even after antibiotics use has ceased. Thus, the negative effects of plasmid carriage can be removed by compensatory adaptation.

1.11 Overall Evaluation of Existing Studies

The ExPEC lineages of the phylogroup B2 are proportionally distributed in many human extra-intestinal diseases, with prevalence in hospital and community acquired ExPEC diseases. They are also commonly encountered in the gut. However, there has been many significant studies on the highly multi-drug resistance lineage, ST131 than its equally important drug susceptible counterparts, ST73 and ST95 lineages. Limited information exists to explain the rationale why these latter lineages though known to be largely susceptible to antibiotics, are also major pathogenic extra-intestinal *E. coli* that are extremely pandemic. Reports have suggested that, though *E. coli* strains of the phylogroup B2, that are antibiotics resistance as well as those that are drug susceptible continue to cause pandemics, none of them has been able to become a dominant cause of extra-intestinal infection (Kallonen *et al*., 2017; Riley *et al*., 2014). This seems to further suggest that antibiotic resistance may not be the only cause of the pandemicity, and pathogenicity observed in the strains of these lineages. However, much of the antimicrobial resistance in *E. coli* is plasmid mediated. Therefore, understanding the characteristics of the plasmids found in these STs in relation to their evolution and dissemination may be the key factor in unveiling what goes on within the genomes of these lineages.

1.12 Research Aims

Over the years, the frequency of AR human pathogenic bacteria has significantly increased around the globe. Several studies have shown that many of these resistant strains are associated with the family Enterobacteriaceae (Skov and Monnet, 2016; Liu *et al*., 2016; Alhashash *et al*., 2016; Nicolas-Chanoine *et al*., 2014; Zhang *et al*.,

2014; Chen *et al*., 2014; Petty *et al*., 2014; Price *et al*., 2013; Golding *et al*., 2012; Dias *et al*., 2009; Coque *et al*., 2008; Nicolas-Chanoine *et al*., 2008). This group of bacteria are frequently recovered from intestinal and extra-intestinal infections. Within the Enterobacteriaceae, *E. coli* is known to be responsible for most of the extra-intestinal infections.

E. coli is sub-divided into phylogroups with major representatives: A, B1, B2 and D. Phylogroup B2 is the major class of *E. coli* responsible for many ExPEC diseases (Gordon *et al*., 2015; Lau *et al*., 2008). Globally, phylogroup B2 strains account for about 70% of ExPEC infections. Within this phylogroup, are three key lineages responsible for most of these ExPEC infections: sequence types (STs) 131, 73 and 95. Many reports have shown that each of these three lineages play an equal role in these infections, with each representing about 10% of *E. co*li recovered from humans (Forde *et al*., 2019; Gordon *et al*., 2017; Kallonen *et al*., 2017; Stephens *et al*., 2017; Chen *et al*., 2014; Gordon, 2010; Nicolas-Chanoine *et al*., 2008). These STs also represent about 30% of isolates recovered from human faecal samples (Gordon *et al*., 2017).

Given how ubiquitous these lineages are among humans, it would be expected that, in a community setting, the three lineages would be exposed to antimicrobials at about the same rate. However, many studies have demonstrated that their antimicrobial profiles differ: ST131 is noted to be highly resistant to antibiotics (Kallonen *et al.*, 2017; Johnson *et al*., 2016), when compared with the other two lineages, STs 73 and 95. These later lineages have been observed to be largely antibiotics susceptible (Kallonen *et al*., 2017; Stephens *et al*., 2017; Gordon *et al*., 2017). The reasons for the

antibiotic susceptibility of these extra-intestinal *E. coli* strains remains largely unknown. Given the genes conferring resistance to antimicrobials are frequently plasmid borne, in this study we investigated the evolution, population dynamics, function and genomic diversity of the plasmids within these lineages.

The first result chapter of this thesis characterised the differences in the genomic diversity of the plasmids in the phylogroup B2 lineages using various bioinformatics tools. The second result chapter investigated the plasmid-host dynamics by examining the rate of plasmid transfers among the ST lineages using plasmid transfer experiments, assessing the impacts of the plasmids on transfer rates using genome wide association studies, determining how well the plasmids are maintained within the host after several generations using stability experiments and examining the fitness costs of the plasmids carriage within the host over time using competition experiments. The third result chapter examined and compared the evolutionary history existing among the two common plasmid type in the three lineages using genome alignments techniques and phylogenetic analysis. The final chapter discussed the findings of this study with reference to the differences in antimicrobial profiles of the three lineages of phylogroup B2, that are equally globally pathogenic and explored the information acquired from this study to improve our understanding of the ecology and evolution of *E. coli* plasmids.

1.13 References

Adams-Sapper, S., Diep, B. A., Perdreau-Remington, F. and Riley, L. W. (2013). Clonal Composition and Community Clustering of Drug-Susceptible and Resistant *Escherichia coli* Isolates from Bloodstream Infections. *Antimicrob Agents Chemother* **57**, 490 – 497.

Adamczyk, M. and Jagura-Burdzy, G. (2003). Spread and survival of promiscuous IncP-1.plasmids. *Acta Biochim Pol* **50**, 425-453.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Molecular.Biology of the Cell. New York: Garland Science. p. G: 35. ISBN 978-0-8153-4072-0.

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. *J Antimicrob Chemother* **71**, 339–43.

Alonso, J. C. (2021). Toxin–Antitoxin Systems in Pathogenic Bacteria. *Toxins* **13**, 74.

Austin, S. and Nordstrom, K. (1990). Partition-Mediated Incompatibility of Bacterial Plasmids. *Cell* **60**, 351–3543.

Austin, S., Ziese, M. and Sternberg, N. (1981). A Novel Role for Site-specific Recombination in Maintenance of Bacterial Replicons. *Cell* **25**, 729–736.

Bailey J. K., Pinyon, J. L., Anantham, S. and Hall, R. M. (2010). Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *J. Med Microbiol* **59**, 1331-1339.

Baltrus, D. A. (2013). Exploring the Costs of Horizontal Gene Transfer. *Trends Ecol Evol* **28**, 489–495.

Banerjee, R., Johnston, B., Lohse, C., Chattopadhyay, S., Tchesnokova, V., Sokurenko, E. V., Johnson, J. R. (2013). The Clonal Distribution and Diversity of Extraintestinal *Escherichia coli* Isolates Vary According to PatientCharacteristics. *Antimicrob Agents Chemother* **57**, 5912–5917.

Barlow, M. (2009). "What Antimicrobial Resistance Has Taught Us about Horizontal Gene Transfer". *Methods in Molecular Biology* **532**, 397–411.

Bengtsson, S., Naseer, U., Sundsfjord, A., Kahlmeter, G. and Sundqvist, M. (2012). Sequence Types and Plasmid Carriage of Uropathogenic *Escherichia coli* Devoid of Phenotypically Detectable Resistance. *J Antimicrob Chemother* **67**, 69–73.

Bennet, P. M., (1995). The Spread of Drug Resistance. In Baumberg, S., Young, J. P. W.,Wellington, E. M. H. and Saunders, J. R. (Eds) Population Genetics of Bacteria.*Cambridge University Press*. Pp317-344.

Bergstrom, C. T., Lipsitch, M. and Levin, B. R. (2000). Natural Selection, Infectious Transfer and the Existence Conditions for Bacterial Plasmids. *Genetics* **155**, 1505–1519.

Bernard, P. and Couturier, M. (1992). Cell killing by the F- Plasmid CcdB Protein Involves Poisoning of DNA-topoisomerase II Complexes. *J Mol Biol* **226**, 735-745.

Bert, F., Johnson, J.R., Ouattara, B., Leflon-Guibout, V., Johnston, B., Marcon, E., Valla, D. Moreau, R. and Nicolas-Chanoine, M.H. (2010). Genetic Diversity and Virulence Profiles of *Escherichia coli* Isolates Causing Spontaneous Bacterial Peritonitis and Bacteraemia in Patients with Cirrhosis*. J Clin Microbiol* **48**, 2709-14.

Bertani, G. and Weigle, J. J. (1953). Host controlled variation in bacterial viruses. *J Bacteriol* **65**,113–121.

Betteridge, T., Partridge, S. R., Iredell, J. R. and Stokes, H. (2011). Genetic context and structural diversity of class 1 integrons from human commensal bacteria in a hospital intensive care unit. *Antimicrob Agents Chemother* **55**, 3939-3943.

Blanco, J., Mora, A., Mamani, R., Lopez, C., Blanco, M., Dahbi, G., Herrera, A., Marzoa, J., Fernandez, V., de la Cruz, F., Martinez-Martinez, L., Alonso, M. P., Nicolas-Chanoine, M. H., Johnson, J. R., Johnston, B., Lopez-Cerero, L., Pascual, A. and Rodriguez-Bano, J. (2013). Four Main Virotypes Among Extended-Spectrum-β-Lactamase-Producing Isolates of *Escherichia coli* O25b:H4-B2-ST131: Bacterial, Epidemiological and Clinical Characteristics. *J Clin Microbiol* **51**, 3358–3367.

Blyton, M. D., Banks, S. C., Peakall, R., Lindenmayer, D. B. and Gordon, D. M. (2014). Not All types of host contacts are equal when it comes to *E. coli* transmission. *Ecol Lett* **17**, 970–978.

Boros, I. Pósfai, G. and Venetianer, P. (1984). "High-copy-number derivatives of the plasmid cloning vector pBR322". *Gene* **30**, 257-60.

Boucher, Y., Cordero, O. X., Takemura, A., Hunt, D. E., Schliep, K., Bapteste, E., Lopez, P., Tarr, C. L. and Polz, M. F. (2011). Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *mBio* **2**, e00335-00310.

Boucher, Y., Labbate, M., Koenig, J.E. and Stokes, H.W. (2007). Integrons: Mobilizable Platforms that Promote Genetic Diversity in Bacteria*. Trends Microbiol **15**,301–309.*

Brantl, S. (2014). Plasmid replication control by antisense RNAs. *Microbiol Spectr* 2:LAS- 0001.

Brantl, S. (2012). Bacterial type I toxin-antitoxin systems. *RNA Biol* **9**, 1488-1490.

Brisse, S., Diancourt, L., Laouénan, C., Vigan, M., Caro, V., Arlet, G., Drieux, L., Leflon-Guibout, V., Mentré, F., Jarlier, V. and Nicolas-Chanoine, M.H. (2012). Phylogenetic Distribution of CTX-M- and Non-Extended-Spectrum-Beta-Lactamase-Producing *Escherichia coli* isolates: group B2 isolates, except clone ST131, rarely produce CTXM Enzymes. *J Clin Microbiol* **50**, 2974-81.

Cabezón, E., Ripoll-Rozada, J., Peña, A., De la Cruz, F. and Arechaga, I. (2015). Towards An Integrated Model of Bacterial Conjugation. *FEMS Microbiol Rev* **39**, 81–95.

Cambray, G., Guerout, A.M. and Mazel, D. (2010*).* Integrons*. Annu Rev Genet **44**, 141–166.*

Cantón, R. and Ruiz-Garbajosa, P. (2011). Co-resistance: An Opportunity for the Bacteria And Resistance Genes. *Curr Opin Pharmacol* **11**, 477–485.

Carattoli, A. (2009). Resistance Plasmid Families in Enterobacteriaceae. *Antimicrob Agents Chemother* **53**, 2227–38.

Carattoli, A. (2013). Plasmids and the spread of resistance. *Int J Med Microbiol* **303**, 298-304

Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K. L. and Threlfall, E. J. (2005). Identification of plasmids by PCR-based replicon typing. *J Microbiol Methods* **63**, 219-28.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Aarestrup, F. M. and Hasman, H. (2014). *In Silico* Detection and Typing of Plasmids Using Plasmidfinder and Plasmid Multilocus Sequence Typing. *Antimicrob Agents Chemother* **58**, 3895–3903.

Chandler, M., Clerget, M. and Galas, D. J. (1982). The Transposition Frequency of ISI-Flanked Transposons is a function of their size. *J Mol Biol* **154**, 221-243.

Chattoraj, D. K. (2000). Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol Microbiol* **37**, 467–476.

Chen, L., Hu, H., Chavda, K. D., Zhao, S., Liu, R., Liang, H., Zhang, W.,Wang, X., Jacobs, M. R., Bonomo, R. A., Kreiswirth, B. N. (2014). Complete Sequence of a KPC-Producing IncN Multidrug-Resistant Plasmid from an Epidemic *Escherichia coli* Sequence Type 131 Strain in China. *Antimicrob Agents Chemother* **58***,* 2422–2425.

Christopher, J. H., Robert, A. M. and Ruth, M. H. (2014). Movement of IS*26*-Associated Antibiotic Resistance Genes Occurs via a Translocatable Unit That Includes a Single IS*26* and Preferentially Inserts Adjacent to Another IS*26. mBio*. **5**, pp 5.

Coque, T. M., Novais, A., Carattoli, A., Poirel, L., Pitout, J., Peixe, L., Baquero, F. and Canton, R., Nordmann, P. (2008). Dissemination of Clonally Related *Escherichia coli* Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15. *Emerg Infect Dis* **14**, 195–200.

Dahbi, G., Mora, A., López, C., Alonso, M. P., Mamani, R., Marzoa, J., Coira, A., García-Garrote, F., Pita, J. M., Velasco, D., Herrera, A., Viso, S., Blanco, J. E., Blanco, M. and Blanco, J. (2013). Emergence of New Variants of ST131 Clonal Group among Extra-intestinal Pathogenic *Escherichia coli* Producing Extended-Spectrum β- Lactamases. *Int J Antimicrob Agents* **42**, 347–351.

Dahlberg, C. and Chao, L. (2003). Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12. *Genetics* **165**, 1641-9.

Dale, A. P. and Woodford, N. (2015). Extra-intestinal Pathogenic *Escherichia coli* (ExPEC): Disease, Carriage and Clones. *J Infect* **71**, 615–626.

Datta, N. and Hedges, R. W. (1973). R factors of compatibility group A. *J Gen Microbiol* **74**, 335–6.

Dobrindt, U., Chowdary, M. G., Krumbholz, G. and Hacker, J. (2010). Genome dynamics and its impact on evolution of *Escherichia coli*. *Medical microbiology and immunology* **199,** 145-154.

De la Cruz, F., Frost, L. S., Meyer, R. J. and Zechner, E. *(*2010*).* Conjugative DNA metabolism in Gram-Negative Bacteria. *FEMS Microbiol Rev **34,** 18-40.*

Derbyshire, K. M. and Gray, T. A. (2014). Distributive Conjugal Transfer: New Insights into Horizontal Gene Transfer and Genetic Exchange in Mycobacteria. *Microbiol Spectr* **2**, MGM2-0022.

Dias, R. C. S., Marangoni, D. V. Smith, S. P. Alves, E. M. Pellegrino, F. L. P. C. Riley, L. W. and Moreira, B. M. ***(***2009*).* Clonal Composition of *Escherichia Coli* Causing Community-acquired Urinary Tract Infections in the State of Rio de Janeiro, *Brazil. Microb Drug Resist* **15,** 303-308 *Microbiol Spectr* **2**, 04.

Di Luca, M. C., Sørum, V., Starikova, I., Kloos, J., Hülter, N., Naseer, U., Johnsen, P. J. and Samuelsen, O. (2017). Low biological cost of carbapenemase-encoding plasmids following transfer from *Klebsiella pneumoniae* to *Escherichia coli*. *Journal of Antimicrobial Chemotherapy* **72**, 85–89.

Dimitriu, T., Misevic, D., Lotton, C., Brown, S. P., Lindner, A. B. and Taddei, F. (2016). Indirect fitness benefits enable the spread of host genes promoting costly transfer of beneficial plasmids. *PLoS Biol* **14**, e1002478.

Dionisio, F., Conceição, I. C., Marques, A. C., Fernandes, L. and Gordo, I. (2005). The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biology letters* **1**, 250–252.

Djordjevic, S. P., Stokes, H. W. and Chowdhury, P. R. (2013). Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Front Microbiol* **4**, 86.

Domingues, S., da Silva, G. J., and Nielsen, K. M. (2012). Integrons: Vehicles and pathways for horizontal dissemination in bacteria. *Mobile genetic elements* **2**, 211–223.

Domingues, S., Harms, K., Fricke, W. F., Johnsen, P. J., da Silva, G. J. and Nielsen, K. M. (2012). Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS Pathog* **8:**e1002837.

Doyle, M., Fookes, M., Ivens, A., Mangan, M. W., Wain, J. and Dorman, C. J. (2007). An H-NS-like stealth protein aids horizontal DNA transmission in Bacteria. *Science* **315**, 251–252.

Ebersbach, G. and Gerdes, K. (2005). Plasmid Segregation Mechanisms. *Annu Rev Genet* **39**, 453–479.

Ebersbach, G., Sherratt, D. J. and Gerdes, K. (2005) Partition-Associated Incompatibility Caused by Random Assortment of Pure Plasmid Clusters. *Mol Microbiol* **56**, 1430–1440.

Escobar-Paramo, P., Grenet, K., Le Menac'h, A., Rode, L., Salgado, E., Amorin, C., Gouriou, S., Picard, B., Rahimy, M. C., Andremont, A., Denamur, E. and Raymond, R. (2004). Large-scale Population Structure of Human Commensal *Escherichia coli* isolates. *Appl Environ Microbiol* **70**, 5698–5700.

Escobar-Paramo, P., Le Menac'h, A., Le Gall, T., Amorin, C., Gouriou, S., Picard, B., Skurnik, D. and Denamur, E. (2006). Identification of Forces Shaping the Commensal *Escherichia coli* Genetic Structure by Comparing Animal and Human Isolates. *Environ Microbiol Rep* **8**, 1975–1984.

Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C. and de la Cruz, F. (2004). A Classification Scheme for Mobilization Regions of Bacterial Plasmids. *FEMS Microbiology Reviews* **28**, 79–100.

Francino, M. P., ed. (2012). Horizontal Gene Transfer in Microorganisms. Caister Academic Press. ISBN 978-1-908230-10-2.

Fratamico, P. M., DebRoy, C., Liu, Y., Needleman, D. S., Baranzoni, G. M. and Feng, P. (2016). Advances in Molecular Serotyping and Subtyping of *Escherichia coli*. *Front Microbiol* **3**, 7:644.

Friehs, K. (2004). Plasmid copy number and plasmid stability. *Adv Biochem Eng Biotechnol* 86, 47-82.

Frost, L. S. and Koraimann, G. (2010). Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol* **5**,1057–1071.

Forde, B.M., Roberts, L.W., Phan, M. D. Peters, K. M*.,* Fleming, B. A*.,* Russell, C. W*.,* Lenherr, S. M., Myers, J. B., Barker, A. P., Fisher, M. A., Chong, T. M., Yin, W. F*.,* Chan, K. G., Schembri, M. A., Mulvey, M. A. and Beatson, S. *A.* (2019) Population dynamics of an *Escherichia coli* ST131 lineage during recurrent urinary tract infection. *Nat Commun* **10,** 3643.

Fozo, E. M., Hemm, M. R. and Storz, G. (2008a). "Small Toxic Proteins and the Antisense RNAs that Repress them". *Microbiol and Molecul Biol Rev* **72**, 579–89.

Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y., Mendieta, K. S., Jones, K. L., Ocampo, A., Rudd, K. E. and Storz, G. (2008b). "Repression of Small Toxic Protein Synthesis by the Sib and OhsC Small RNAs". *Molecul Microbiol* **70**, 1076–93.

Fozo, E. M., Makarova, K. S., Shabalina, S. A., Yutin, N., Koonin, E. V. and Storz, G. (2010)."Abundance of Type I Toxin-antitoxin Systems in Bacteria: Searches for new. Candidates and Discovery of Novel Families". *Nucleic Acids Research* **38**, 3743–59.

Garcillán-Barcia, M. P., Alvarado, A. and de la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol Rev* **35**, 936–956.

Garcillán-Barcia, M. P., Francia, M. V. and de la Cruz, F. (2009). The Diversity of Conjugative Relaxases and its application in Plasmid Classification. *FEMS Microbiol Rev* **33**, 657–687.

Gerdes, K., Bech, F. W., Jørgensen, S. T., Løbner-Olesen, A., Rasmussen, P. B., Atlung, T., Boe, L., Karlstrom, O., Molin, S., and von Meyenburg, K. (1986). Mechanism of Postsegregational Killing by the Hok Gene Product of the ParB System of Plasmid R1 and its Homology with the RelF Gene Product of the *E. coli* RelB Operon. *The EMBO journal* **5**, 2023–2029.

Gerdes, K., Gultyaev, A. P., Franch, T., Pedersen, K. and Mikkelsen, N. D. (1997). Antisense RNA-Regulated Programmed Cell Death. Annu. *Rev Genet* **31**, 1–31.

Gerdes, K., Moller-Jensen, J. and Bugge -Jensen, R. (2000). Plasmid and Chromosome Partitioning: Surprises from Phylogeny. *Mol Microbiol* **37**, 455–466.

Gerdes, K., Rasmussen, P. B. and Molin, S. (1986). Unique Type of Plasmid Maintenance Function: Postsegregational Killing of Plasmid-Free Cells. *Proc Natl Acad Sci* **83**, 3116–3120.

Gerdes, K. and Wagner, E. G. (2007). RNA Antitoxins. *Curr Opin Microbiol* **10**, 117–124.

Gibreel, T. M., Dodgson, A. R., Cheesbrough, J., Fox, A. J., Bolton, F. J. and Upton, M. (2012). Population Structure, Virulence Potential and Antibiotic Susceptibility of Uropathogenic *Escherichia coli* from Northwest England. *J Antimicrob Chemother* 67, 346–356.

Giedraitiene, A., Vitkauskiene, A., Pavilonis, A., Patamsyte, V., Genel, N., Decre, D. and Arlet, G. (2017). Prevalence of O25b-ST131 Clone among *Escherichia coli* Strains Producing CTX-M-15, CTX-M-14 and CTX-M-92 beta-lactamases. *Infect Dis* **49**,106–112.

Gillings, M.R. (2014). Integrons: Past, Present and Future. *Microbiol Mol Biol Rev* **78**, 257-277.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M. and Stokes, H. W. (2008). The Evolution of Class 1 Integrons and the Rise of Antibiotic Resistance*. Journal of bacteriology **190***, 5095–5100.

Golding, G. R., Persaud, N., Levett, P. N., McDonald, R. R., Irvine, J., Nsungu, M., Woods, S., Khan, M., Mataseje, L. F. and Mulvey, M. R. (2012). Characterization of *Escherichia coli* Urinary Tract Infection Isolates in Remote Northern Saskatchewan Communities: the Northern Antibiotic Resistance Partnership. *Diagn Microbiol Infect Dis* **74**, 242–247.

Gordon, D. M. (2010). Strain typing and the ecological structure of *Escherichia coli*. *J. AOAC International* **93**, 974–984.

Gordon, D. M., Bauer, S. and James R. Johnson, J. R. (2002). The genetic structure of *Escherichia coli* populations in primary and secondary habitats*. Microbiology*. **148**, 1513–1522.

Gordon, D. M., Geyik, S., Clermont, O., O'Brien, C. L., Huang, S., Abayasekara, C., Rajesh, A., Kennedy, K., Collignon, P., Pavli, P., Rodriguez, C., Johnston, B. D., Johnson, J. R., Decousser, J. W. and Denamur, E. (2017). Fine-Scale Structure Analysis Shows Epidemic Patterns of Clonal Complex 95, a Cosmopolitan Escherichia coli Lineage Responsible for Extraintestinal Infection. *mSphere*. pp 2.

Gordon, D. M., O'Brien, C. L. and Pavli, P. (2015). *Escherichia coli* Diversity in the Lower Intestinal Tract of Humans. *Environ. Microbio. Reports*. **7**, 642–648.

Gordon, D. M., Stern, S. E. and Colington, P. J. (2005). Influence of the Age and Sex of Human Hosts on the Distribution of *Escherichia coli* ECOR Groups and Virulence Traits. *Microbiology* **151**, 15–23.

Gordon, G. S. and Wright, A. (2000). DNA Segregation in Bacteria. Annu Rev Microbiol **54**, 681–708.

Gomis-Rüth, F. X. and Coll, M. (2006). Cut and Move: Protein Machinery for DNA Processing in Bacterial Conjugation. *Curr Opin Struct Biol* **16**, 744–52.

Griffiths, A.J.F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. and Gelbar, W. M. (2000). An Introduction to Genetic Analysis. 7th edition. New York: W. H. Freeman. ISBN-10: 0-7167-3520-2**.**

Guglielmini, J., Néron, B., Abby, S. S., Garcillán-Barcia, M. P., de la Cruz, F. and Rocha, E.P.C. (2014). Key components of the eight classes of type IV secretion Systems involved in Bacterial conjugation or protein secretion. *Nucleic Acids Res* **42**, 5715–5727.

Gyles, C. and Boerlin, P. (2014). "Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease". Veterinary Pathology 51, 328–40.

Haase, J., Lurz, R., Grahn, A. M., Bamford, D. H. and Lanka, E. (1995). Bacterial Conjugation Mediated by Plasmid RP4: RSF1010 Mobilization, Donor-Specific Phage Propagation and Pilus Production Require the Same Tra2 Core Components Of a Proposed DNA Transport Complex. *J Bacteriol* **177**, 4779–91.

Harms, A., Brodersen, D. E., Mitarai, N. and Gerdes, K. (2018). Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology *Mol Cell* **70**, 768-784.

Hall, R. M. and Collis, C. M. (1995). "Mobile Gene Cassettes and Integrons*:* Capture and Spread of Genes by Site-Specific Recombination"*. Molecular Microbiology.* **15***,* 593–*600.*

Hall, R. M. and Stokes, H. (2004). Integrons or super integrons? *Microbiology* **150**, 3-4.

Hawkey, P. M. and Jones, A. M. (2009). "The Changing Epidemiology of Resistance". *The Journal of Antimicrobial Chemotherapy* **64**, 3-10.

Hayes, F. (2003). Toxins-antitoxins: Plasmid Maintenance, Programmed Cell Death and Cell Cycle Arrest. *Science* **301**, 1496–1499.

Hayes*,* F. and Melderen, L*. V.* (2011*).* Toxins-antitoxins: diversity, evolution and function. *Critical Reviews in Biochemistry and Molecular Biology, **46**, 386 - 408.*

Horner, C., Fawley, W., Morris, K., Parnell, P., Denton, M. and Wilcox, M. (2014). *Escherichia coli* Bacteraemia: 2 Years of Prospective Regional Surveillance (2010–12). *J Antimicrob Chemother* **69,** 91-100.

Hotopp, D. J. C. (2011). "Horizontal Gene Transfer Between Bacteria and Animals". *Trends In Genetics* **27**, 157–63.

Jacquier, H., Zaoui, C., Sanson-le Pors, M. J., Mazel, D. and Berçot, B. (2009). Translation regulation of integrons gene cassette expression by the attC sites. *Mol Microbiol* **72**, 1475-1486. Jang, J., Hur, H. G., Sadowsky, M., Byappanahalli, M., Yan, T. and Ishii, S. (2017). Environmental *Escherichia coli*: Ecology and Public Health Implications-A Review*. J Appl Microbiol*, **123**, 570-581.

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., Carbonnelle, E., Lortholary, O., Clermont, O., Denamur, E., Picard, B., Nassif, X. and Brisse, S. (2008). Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* **9**, 560.

Johnson, J. R., Brian, J., Connie, C., Kuskowski, M. A. and Mariana, C. (2010). *Escherichia coli* Sequence Type ST131 as the Major Cause of Serious Multidrug-Resistant *E. col*i Infections in the United States. *Clin Infect Dis* **51**, 286–294.

Johnson, T. J., Danzeisen, J. L., Youmans, B., Case, K., Llop, K., Munoz-Aguayo, J., Flores-Figueroa, C., Aziz, M., Stoesser, N., Sokurenko, E., Price, L. B. and Johnson, J. R. (2016). Separate F-Type Plasmids Have Shaped the Evolution of the *H*30 Subclone of *Escherichia coli* Sequence Type 131. *Clinical Science and Epidemiology mSphere,* **1** (4).

Johnson, C. M. and Grossman, A. D. (2015). "Integrative and Conjugative Elements (ICEs): What They Do and How They Work". *Annual Review of Genetics* **42**, 577-601.

Johnson, J. R., Johnston, B., Kuskowski, M. A., Nougayrede, J. P., and Oswald, E. (2008). Molecular Epidemiology and Phylogenetic Distribution of the *Escherichia coli* Pks Genomic Island. *Journal of clinical microbiology*, **46**, 3906–3911.

Johnson, J. R., Owens, K. L., Clabots, C. R., Weissman, S. J. and Cannon, S. B. (2006). Phylogenetic relationships among clonal groups of extraintestinal pathogenic *Escherichia coli* as assessed by multi-locus sequence analysis. *Microbes Infect* **8,** 1702–1713.

Johnson, T. J., Singer, R. S., Isaacson, R. E., Danzeisen, J. L., Lang, K., Kobluk, K., Rivet, B., Borewicz, K., Frye, J. G., Englen, M., Anderson, J., and Davies, P. R. (2015). In Vivo Transmission of an IncA/C Plasmid in *Escherichia coli* Depends on Tetracycline Concentration, and Acquisition of the Plasmid

Results in a Variable Cost of Fitness. *Applied and environmental microbiology*, **81**, 3561–3570.

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 J Infect Dis* **207**, 919–928.

Johnston, C., Martin, B., Fichant, G., Polard, P. and Claverys, J. P. (2014). "Bacterial Transformation: Distribution, Shared Mechanisms and Divergent Control". *Nature Reviews Microbiology* **12**, 181–96.

Kallonen, T., Brodrick, H. J, Harris, S. R, Coriander, 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**, 1437-1449.

Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004). Pathogenic *Escherichia coli. Nat Rev Microbiol* **2**, 123–140.

Kay, E., Vogel, T. M., Bertolla, F., Nalin, R. and Simonet, P. (2002). "In Situ Transfer of Antibiotic Resistance Genes from Transgenic (Transplastomic) Tobacco Plants to Bacteria". *Applied and Environmental Microbiology* **68**, 3345–51.

Keeling, P. J., Palmer, J. D. (2008). "Horizontal Gene Transfer in Eukaryotic Evolution". *Nature Reviews Genetics* **9**, 605–18.

Khosravi, M., Ryan, W., Webster, D. A. and Stark, B. A. (1990). Variation of Oxygen Requirement with Plasmid Size in Recombinant *Escherichia coli*. *Plasmid* **23,** 138-143.

Kline, B. C. (1985). A review of mini-F plasmid maintenance. *Plasmid* **14**, 1-16.

Koonin, E. V., Makarova, K. S. and Aravind, L. (2001). "Horizontal Gene Transfer in Prokaryotes: Quantification and Classification". *Annual Review of Microbiology* **55**, 709–42.

Kovalevskaya, N. P. *(2002).* "Mobile Gene Cassettes and Integrons*". Molecular Biology. **36**, 196–201.*

Lau, S. H., Kaufmann, M. E., Livermore, D. M., Woodford, N., Willshaw, G. A., Cheasty, T., Stamper, K., Reddy, S., Cheesbrough, J., Bolton, F. J., Fox, A. J. and Upton, M. (2008). UK Epidemic *Escherichia coli* Strains A-E, with CTX-M-15 Beta-Lactamase, all belong to the International O25:H4-ST131 Clone. *J Antimicrob Chemother* **62**, 1241-1244.

Lau, S. H., Reddy, S., Cheesbrough, J., Bolton, F. J., Willshaw, G., Cheasty, T., Fox, A. J And Upton, M. (2008). Major uropathogenic *Escherichia coli* strain isolated in the northwest of England identified by multilocus sequence typing. *J Clin Microbiol* **46**, 1076-80.

Lawley, T. D., Klimke, W. A., Gubbins, M. J. and Frost, L. S. (2003). F- Factor Conjugation is a True Type IV Secretion System. *FEMS Microbiol Lett* **224**, 1–15.

Lee, S. W. and Edlin, G. (1985). Expression of tetracycline resistance in pBR322 derivatives reduces the reproductive fitness of plasmid-containing *Escherichia coli*. *Gene* **39**, 173-180.

Leimbach, A., Hacker, J. and Dobrindt, U. (2013). *E. coli* as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity. *In:* Dobrindt, U., Hacker, J. H. and Svanborg, C. (eds.) *Between Pathogenicity and Commensalism.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Lenski, R. E. and Bouma, J. A. (1987). Effects of Segregation and Selection on Instability of Plasmid pACYC184 in *Escherichia coli* B. Jour of bacteriol **169**, 5314-5316.

Levin, B. R. (1995). Conditions for the Evolution of Multiple Antibiotic Resistance Plasmids: A Theoretical and Experimental Excursion. *Population Genetics of Bacteria*, Pp.175–192. Edited by S. Baumberg. Cambridge/New York: Cambridge University Press.

Lewis-Kittell, B. and Helinski, D. R. (1993). Plasmid Incompatibility and Replication Control. *Bacterial Conjugation.* Clewell, D. B. (Ed). Plenum Press, New York. Pp 223- *237.*

Li, Z., Lin, Y., Lu, L., Wang, K., Yang, L., Li, P., Li, J., Jia, L., Li, P. and Song, H. (2020) Genetic Characterisation of a Complex Class 1 Integron in an NDM-1-Producing *Citrobacter freundii* ST396 Clinical Strain Isolated from a Urine Sample. *Jour Glob Antimicro Resist* **23**, 64-66.

Lili, L. N., Britton, N. F. and Feil, E. J. (2007). The Persistence of Parasitic Plasmids. *Genetics* **177**, 399–405.

Liu, H., Wang, H., Huang, M., Mei, Y., Gu, B., Wu, R., Huang, Y., Chen, Y., Xu, Y. and Wang T. (2013). Analysis of antimicrobial resistance and class 1 integrons among strains from upper respiratory tract of healthy adults. *J Thorac Dis* **5**, 149.

Liu, Y. Y., Wang, Y., Walsh, T. R., Yi, L. X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L. F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J. H. and Shen, J. (2016). Emergence of Plasmid-Mediated Colistin Resistance Mechanism MCR-1 in Animals and Human Beings in

China: a Microbiological and Molecular Biological Study. *Lancet Infect Dis* **16**, 161-168.

Llosa, M., Gomis-Rüth, F. X., Coll, M. and De la Cruz, F. d. (2002). Bacterial conjugation: A Two-Step Mechanism for DNA Transport. *Mol Microbiol* **45**, 1–8.

Lukjancenko, O., Wassenaar, T. M. and Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microbial ecology* **60,** 708-720.

Ma, Z., Geng, J., Yi, L., Xu, B., Jia, R., Li, Y., Meng, Q., Fan, H. and Hu, S. (2013). Insight Into the Specific Virulence Related Genes and Toxin–antitoxin Virulent Pathogenicity Islands in Swine Streptococcosis Pathogen *Streptococcus Equis* sp. Zooepidemicus Strain ATCC 35246. *BMC Genomics* **14**, 377.

Madigan, M. and Martinko, J. (editors) (2006). Brock Biology of Microorganisms (11th Ed.). Prentice Hall. ISBN 978-0-13-144329-7.

Magnuson, R. D. (2007). Hypothetical Functions of Toxin-Antitoxin Systems. *Journal of Bacteriology* **189**, 6089-6092.

Malamy, M. H., Fiandt, M. and Szybalski, W. (1972) Electron Microscopy of Polar Insertions In the Lac Operon of *Escherichia coli Mol Gen Genet* **119**, 207–222.

Malys, N. (2012). "Shine-Dalgarno Sequence of Bacteriophage T4: GAGG Prevails in Early Genes". *Molecular Biology Reports* **39**, 33–9.

Manges, A. R., Harel, J., Masson, L., Edens, T. J., Portt, A., Reid-Smith, R. J., Zhanel, G .G., Kropinski, A. M. and Boerlin, P. *(*2015*).* Multilocus Sequence Typing And Virulence Gene Profiles Associated with *Escherichia coli* from Human And Animal Sources*. Foodborne Pathog Dis* **12***,* 302–310.

Marraffini, L. A., Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science* **322**, 1843–1845.

Martinez-Freijo**,** P., Fluit, A. C., Schmitz**, F. J., Gre**k, V. S**.,** Verhoef, J. and **J**ones, M. E**. (**1998). Class I Integrons in Gram-Negative Isolates from Different European Hospitals and Association with Decreased Susceptibility to Multiple Antibiotic Compounds. J. *Antimicrob Chemother* **42,** 689-696.

Martinez-Medina, M., Mora, A., Blanco, M., López, C., Alonso, M. P., Bonacorsi , S., M. H., Darfeuille-Michaud, A., Garcia-Gil, J. and Blanco, J. (2009). Similarity and Divergence among Adherent-Invasive *Escherichia coli* and Extraintestinal Pathogenic *E. coli* strains*. J Clin Microbiol* **47**, 3968–3979.

Matsumura, Y. Yamamoto, M., Nagao, M., Hotta, G., Matsushima, A., Ito, Y., Takakura, S. and Ichiyama, S. (2012)*.* Emergence and spread of B2-ST131-O25b, B2-ST131-O16 and D-ST405 Clonal Groups Among Extended-

Spectrum-Beta-Lactamase-Producing *Escherichia coli* in Japan. J *Antimicrob Chemother* **67**, 2612-2620.

Mazel, D. (2006). "Integrons: Agents of Bacterial Evolution". *Nature Reviews Microbiology.* **4,** 608–620.

McClintock, B. (1950) The origin and Behaviour of Mutable Loci in Maize. *Proc Natl Acad Sci* **36**, 344–355.

McGivern, B. B., McDonell, R. K., Morris, S. K., LaPara, T. M. and Donato, J. J. (2021) Novel Class 1 Integron Harboring Antibiotic Resistance Genes in Wastewater-Derived Bacteria as Revealed by Functional Metagenomics. *Plasmid* **114**, 102563.

Merino, I., Shaw, E., Horcajada, J. P., Cercenado, E., Mirelis, B., Pallarés, M. A., Gómez, j., Xercavins, M., Martínez-Martínez, L., De Cueto, M., Cantón, R., Ruiz-Garbajosa, P., Bunshow, E., Sánchez-Carrillo, C., Padilla, B., Benito, N., Gamallo, R., Riera, M., Calbo, E., Fariñas, M. C., Gonzalo, M., Pascual, A., Rodríguez-Baño, J. and Pintado, V. (2016). CTX-M-15-H30Rx-ST131 Subclone is one of the main causes of Healthcare-Associated ESBL-Producing *Escherichia coli* Bacteraemia of Urinary origin in Spain. *J Antimicrob Chemother* **71**, 2125–2130.

Michael, C. A. and Labbate, M. (2010). Gene cassette transcription in a large integron-associated array. *BMC Genet* **11**, 82.

Million-Weaver, S. and Camps, M. (2014). Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid* **75**, 27-36.

Mora**,** A., Lopez, C., Dabhi, G., Blanco, M., Blanco, J. E., Alonso, M. P., Herrera, A., Mamani, R., Bonacorsi, S., Moulin-Schouleur, M. and Blanco, J. (2009). Extraintestinal Pathogenic *Escherichia coli* O1:K1:H7/NM from Human And Avian Origin: Detection of Clonal Groups B2 ST95 and D ST59 with Different Host Distribution. *BMC Microbiol* **9**, 132.

Muñoz-López, M. and García-Pérez, J. L. (2010). "DNA Transposons: Nature and Applications In Genomics. *Current Genomics* **11**, 115–128.

Mutai, W. C., Waiyaki, P. G., Kariuki, S., Kariuki, S. and Muigai, A. W. T. (2019). Plasmid Profiling and Incompatibility Grouping of Multidrug Resistant *Salmonella enterica* serovar Typhi Isolates in Nairobi, Kenya. *BMC Res Notes* **12,** 422.

Nandi, S., Maurer, J. J., Hofacre, C. and Summers, A. O. (2004). Gram-Positive Bacteria are Major Reservoir of Class 1 Antibiotic Resistance Integrons in Poultry Litter. *Proc Natl Acad Sci* 101, 7118-22.

Naas, T., Mikami, Y., Imai, T., Poirel, L. and Nordmann, P*. (2001).* Characterization of In53, a Class 1 Plasmid and Composite Transposon-located Integron of

Escherichia coli which carries an unusual array of Gene Cassettes*. J. Bacteriol. **183**, 235–249.*

Nicolas-Chanoine, M. H., Bertrand, X. and Madec, J. Y. (2014). *Escherichia coli* ST131, an Intriguing Clonal Group. *Clin Microbiol Rev* **27**, 543–574.

Nicolas-Chanoine, M. H., Blanco, J., Leflon-Guibout, V., Demarty, R., Alonso, M. P., Caniça, M. M., Park, Y. Y., Lavigne, J. P., Pitout, J. and Johnson, J. R. (2008). Intercontinental Emergence of *Escherichia coli* Clone O25:H4-ST131 Producing CTX-M-15. *J Antimicrob Chemother* **61**, 273-281.

Nielsen, K. M. (1998). "Barriers to Horizontal Gene Transfer by Natural Transformation in Soil Bacteria". *APMIS* **84**, 77–84.

Noack, D., Roth, M., Geuther, R., Muller, G., Undisz, K., Hoffmeier, C. and Gaspar, S. (1981). Maintenance and genetic stability of vector plasmids pBR322 and pBR325 in *Escherichia coli* K12 strains grown in a chemostat. *Mol Gen Genet* 184, 121-124.

Nordström, K. and Austin, S. J. (1989). Mechanisms that Contribute to the Stable Segregation of Plasmids. *Annu Rev Genet* **23,** 37-69.

Novick, R. P. (1987). Plasmid Incompatibility. *Microbiol Rev* **51**, 381–395.

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000). "Lateral Gene Transfer and the Nature of Bacterial Innovation". *Nature* **405**, 299–304.

Olesen, B., Hansen, D. S., Nilsson, F., Frimodt-Møller, J., Leihof, R. F., Struve, C., Scheutz, F., Johnston, B., Krogfelt, K. A. and Johnson, J. R. (2013). Prevalence and Characteristics of the Epidemic Multiresistant *Escherichia coli* ST131 Clonal Group Among Extended-Spectrum Beta-Lactamase-Producing *E. coli Isolates* in Copenhagen, Denmark. *J Clin Microbiol* **51**, 1779–1785.

Oliveira, P. H., Touchon, M., Cury, J. and Rocha, E. P. (2017). "The Chromosomal Organization of Horizontal Gene Transfer in Bacteria". *Nature Communications* **8**, 841.

Organisation for Economic Co-operation and Development (2010). Safety Assessment of Transgenic Organisms, Volume 4: OECD *Consensus Documents*, pp.171- 174

Partridge, S. R. (2011*). Analysis of Antibiotic Resistance Regions in Gram-Negative Bacteria*. *FEMS Microbiol Rev **35**, 820–855.*

Partridge, S. R., Tsafnat, G., Coiera, E. and Iredell, J. R. (2009*).* Gene Cassettes and Cassette Arrays in Mobile Resistance Integrons*. FEMS Microbiol Rev **33**,* 757–784*.*

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-Baño, J., Pascual, A., Pitout, J. 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. *Proc Natl Acad Sci* **111**, 5694-9.

Popov, M., Petrov, S., Kirilov, K., Nacheva, G. and Ivanov, I. (2009) Segregational Instability in *E. Coli* of Expression Plasmids Carrying Human Interferon Gamma Gene and its 3'- End Truncated Variants. *Biotechnology and Biotechnological Equipment* **23**, 840-843.

Porse, A., Schønning, K., Munck, C. and Morten O.A. Sommer, M. O. (2016). Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Molecular Biology and Evolution* **33**, 2860–2873.

Prentice, M. B., James, K. D., Parkhill, J., Baker, S. G., Stevens, K., Simmonds, M. N., Mungall, K. L., Churcher, C., Oyston, P, C., ZTitball, R. W., Wren, B. W., Wain, J., Pickard, D., Hien, T. T., Farrar, J. J. and Dougan, G. (2001). *Yersinia pestis* pFra Shows Biovar-Specific Differences and Recent Common Ancestry with a *Salmonella enterica* Serovar Typhi Plasmid. *J Bacteriol* **183**, 2586-2594.

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-B-Lactamase-Producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30. Rx. *MBio* **4**, 377-13.

Qi, W., Lacher, D., Bumbaugh, A., Hyma, K., Quellette, L., Large, T. and Whittam, T. (2004). EcMLST: an online database for multi locus sequence typing of pathogenic *escherichia coli*. *Comput Syst Bioinformatics Conf* 520–521.

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V. and Ravel, J. (2008). The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. *Journal of Bacteriology* **190,** 6881-6893.

Renner, S. S. and Bellot, S. (2012). "Horizontal Gene Transfer in Eukaryotes: Fungi-to-Plant and Plant-to-Plant Transfers of Organellar DNA". Genomics of Chloroplasts and Mitochondria. Advances in Photosynthesis and Respiration. Springer Science/Business Media B.V. **35**, 223–235.

Reyes-Lamothe, R., Tran, T., Meas, D., Lee, L., Li, A. M., Sherratt, D. J., and Tolmasky, M. E. (2014). High-Copy Bacterial Plasmids Diffuse in the Nucleoid-Free Space, Replicate Stochastically and are Randomly Partitioned at Cell Division. *Nucleic acids research*, **42**, 1042–1051.

Riley, L. W. (2014). Pandemic lineages of extraintestinal pathogenic *Escherichia coli. Clinical Microbiology and Infection* **20,** 380-390.

Robinson, K. M., Sieber, K. B. and Hotopp, D. J. C. (2013). "A Review of Bacteria-Animal Lateral Gene Transfer May Inform our Understanding of Diseases like Cancer". *PLOS Genetics* **9** (10): e1003877.

Rosenberg, S. M. and Hastings, P. J. (2001). F Factor, Editor(s): Sydney Brenner, Jefferey, H. Miller, *Encyclopedia of Genetics*, Academic Press, Pp 677-680. ISBN 9780122270802.

Rowe-Magnus, D. A., Guerout, A. M., Biskri, L., Bouige, P. and Mazel, D. (2003). Comparative Analysis of Superintegrons: Engineering Extensive Genetic Diversity in the Vibrionaceae. *Genome Res* **13**, 428–442.

Russo, T. A. and Johnson, J. R. (2003). Medical and economic impact of extraintestinal infections due to *Escherichia coli*: an overlooked epidemic. *Microbes Infect* **5**, 449–456.

Salipante, S. J., Roach, D. J., Kitzman, J. O., Snyder, M. W., Stackhouse, B., Butler-Wu, S. M., Lee, C., Cookson, B. T. and Shendure, J. (2015). Large-Scale Genomic Sequencing of Extraintestinal Pathogenic *Escherichia coli* Strains. *Genome Res* **25**, 119 –128.

Salyers, A. A., Shoemaker, N. B., Stevens, A. M. and Li, L. Y. (1995). Conjugative transposons: An unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* **59**, 579-590.

San Millan, A., Escudero, J. A., Gifford, D. R., Mazel, D. and MacLean, R. C. (2016). Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. *Nat Ecol Evol* **7**, 1-10.

San Millan, A., Heilbron, K. and Maclean, R. C. (2013). Positive Epistasis between Co-Infecting Plasmids Promotes Plasmid Survival in Bacterial Populations. *ISME J **8***, 601–612.

San Millan, A., Heilbron, K. and MacLean, R. (2014a). Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *ISME J* **8,** 601–612.

San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z. V., McLean, A. R., Cooper, B. S. and MacLean, R. C. (2014b). Positive Selection and Compensatory Adaptation Interact to Stabilize Non-Transmissible Plasmids. *Nat Commun* **5**, 5208.

Schröder, G. and Lanka, E. (2005). The Mating Pair Formation System of Conjugative Plasmids-A Versatile Secretion Machinery for Transfer of Proteins and DNA. *Plasmid* **54**, 1–25.

Scott, J. R. 1984. Regulation of plasmid replication. *Microbiol Rev* **48**, 1-23.

Seo, J. H. and Bailey, J. E. (1985). Effects of recombinant plasmid content on growth properties and cloned gene product formation in *Escherichia coli*. *Biotechnol. Bioeng* **27**, 1668-1674.

Sepp, E., Stsepetova, J., Lõivukene, K., Truusalu, K., Kõljalg, S., Naaber, P. and Mikelsaar, M. (2009). The occurrence of antimicrobial resistance and class 1 integrons among commensal *Escherichia coli* isolates from infants and elderly persons. *Ann Clin Microbiol Antimicrob* **8,** 34.

Shachrai, I., Zaslaver, A., Alon, U. and Dekel, E. (2010). Cost of Unneeded Proteins in *E. coli* is Reduced after Several Generations in Exponential Growth. *Mol Cell* 38, 758–767.

Sherley, M., Gordon, D. M. and Collignon, P. J. (2003). Species Differences in Plasmid Carriage in the Enterobacteriaceae. Short Communication. *ScienceDirect* **49**, 79-85.

Shi, L., Zheng, M., Xiao, Z., Asakura, M., Su, J., Li, L. and Yamasaki, S. (2006). Unnoticed Spread of Class 1 Integrons in Gram-Positive Clinical Strains Isolated in Guangzhou, China. *Microbiol Immunol* **50**, 463–7.

Shintani, M., Sanchez, Z. K. and Kimbara, K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* **6**, 242.

Skov, R. L. and Monnet, D. L. (2016). Plasmid-Mediated Colistin Resistance (mcr-1 gene): Three Months Later, the Story Unfolds. *Euro Surveill* **21**, 30155.

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. and de la Cruz, *F.* (2010). "Mobility of Plasmids. *Microbiology and Molecular Biology Reviews.* **74***,* 434–452.

Smith, M. A. and Bidochka, M. J. (1998). Bacterial Fitness and Plasmid Loss: The Importance of Culture Conditions and Plasmid Size. *Can J Microbiol* **44**, 351-355.

Smith, J. L., Pina, J. L., Fratamico, M. and Gunther, N. W. (2007). Extra- intestinal Pathogenic *Escherichia coli*. *Foodborne Pathogens and Disease* **4**, 134-163.

Snyder, L., Peters, J. E., Henkin, T. M. and Champness, W. (2013*). "Lysogeny: the λ* Paradigm *and* the Role of Lysogenic Conversion *in* Bacteria*l* Pathogenesis*". Molecular Genetics of Bacteria (4th ed.).* ASM Press. Pp. 340–343*.*

Stalker, B. G., Thomas, C. M. and Helinski, D. R. (1981). Nucleotide Sequence of the Origin of Replication of the Broad-host-range PlasmidRK2. *Mol Gen Genet **181,** 8-12.*

Starikova,I., Al-Haroni, M., Werner, G., Roberts, A. P., Sørum, V., Nielsen, K. M. and Johnsen, P. J. (2013). Fitness costs of various mobile genetic elements in *Enterococcus faecium* and *Enterococcus faecalis*. *Journal of Antimicrobia Chemotherapy* **68**, 2755–2765.

Stearns, S. C., and Hoekstra, R. F. (2005). Evolution: An introduction (2nd ed.). Oxford, NY: Oxford Univ. Press. pp. 38-40.

Stephens, C. M., Adams-Sapper, S. Manraj Sekhon, Johnson, J. R. and Riley, L. W. (2017). Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. *mSphere*, **2** (2). DOI: 10.1128/mSphere.00390-16

Stewart, F. M. and Levin, B. R. (1977). The Population Biology of Bacterial Plasmids: A Priority Conditions for the Existence of Conjugationally Transmitted Factors. Genetics **87**, 209-228.

Stokes, H. W. and Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *Fems Microbiology Reviews* 35**,** 790-819.

Stokes, H. W., Nesbø, C. L., Holley, M., Bahl, M. I., Gillings, M. R., Boucher, Y. (2006). Class 1 integrons potentially predating the association with tn402-like transposition genes are present in a sediment microbial community. *J Bacteriol* **188**, 5722–30.

Strauch, E., Lurz, R. and Beutin, L. (2001). "Characterization of a Shiga Toxin-Encoding Temperate Bacteriophage of *Shigella sonnei*". *Infection and Immunity* **69**, 7588–95.

Summers, D. K., Beton, C. W. and Withers, H. L. (1993). Multicopy Plasmid Instability: the Dimer Catastrophe Hypothesis. *Mol Microbiol* **8**, 1031–1038.

Thomas, C. M. (2014). Plasmid Incompatibility. In: Bell, Ellis, Bond, Judith S., Klinman, Judith P., Masters, Bettie Sue Siler, Wells, Robert D. (Eds.), Part of the “Plasmid Genomes” Section of Molecular Life Sciences: *FEMS Microbiol Rev* **35**, 936-56.

Thomas, C. M. and Summers, D. (2008). Bacterial Plasmids. In eLS, (Ed.). *https://doi.org/10.1002/9780470015902.a0000468.pub2*.

Todar, K. (2012)."Bacterial Resistance to Antibiotics". The Microbial World: Lectures in Microbiology, Department of Bacteriology, University of Wisconsin-Madison.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M. E., Frapy, E., Garry, L., Ghigo, J. M., Gilles, A. M., Johnson, J., Le Bouguénec, C., Lescat, M., Mangenot, S., Martinez-jéhanne, V., Matic, I., Nassif, X., Oztas, S., Petit,

M. A., Pichon, C., Rouy, Z., Ruf, C. S., Schneider, D., Tourret, I., Vacherie, B., Vallenet, D., Médigue, C., Rocha, E. P. C. and Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS genetics* **5,** e1000344-e1000344.

Touchon, M., Perrin, A., de Sousa, J. A. M., Vangchhia, B., Burn, S., o'Brien, C. L., Denamur, E., Gordon, D. and Rocha, E. P. (2020). Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *Biorxiv* 02.12.945709.

Tsang, J. (2017). Bacterial plasmid addiction systems and their implications for antibiotic Drug development. *Postdoc journal : a journal of postdoctoral research and postdoctoral affairs* **5**, 3–9.

Tsuchimoto, S., Nishimura, Y. and Ohtsubo, E. (1992). The Stable Maintenance System Pem of Plasmid R100: Degradation of PemI Protein may allow PemK Protein to Inhibit Cell Growth. *J Bacteriol* **174**, 4205–4211.

Tsukamoto, N., Ohkoshi, Y., Okubo, T., Sato, T., Kuwahara, O., Fujii, N., Tamura, Y. and Yokota, S. (2013). High Prevalence of Cross-Resistance to Aminoglycosides In Fluoroquinolone-Resistant *Escherichia coli* Clinical Isolates*. Chemother* **59**, 379– 384.

Van der Bij, A. K., Peirano, G., Pitondo-Silva, A., Pitout, J. D. D. (2012). The Presence of Genes Encoding for Different Virulence Factors in Clonally Related *Escherichia coli* that Produce CTX-Ms. *Diagn Microbiol Infect Dis* **72,** 297–302.

van Essen-Zandbergen, A., Smith, H., Veldman, K. and Mevius, D. (2007). Occurrence and characteristics of class 1, 2 and 3 integrons in *Escherichia coli*, *Salmonella* and *Campylobacter* spp. in the Netherlands. *J Antimicrob Chemoth* **59, 7**46–750.

Van Essen-Zandbergen, A., Smith, H., Veldman, K. and Mevius, D. (2009) In Vivo Transfer of an incFIB Plasmid Harbouring a Class 1 Integron with Gene Cassettes DfrA1-aadA1. *Vet Microbiol* **137**, 402-7.

Van Melderen, L. (2010).Toxin-antitoxin systems: why so many, what for? *Curr Opinion Microbiol* **13**, 781-785.

Van Melderen, L. and Saavedra De Bast, M. (2009). Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities? *PLoS Genet* **5**, e1000437.

Vogwill, T., MacLean, R. C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol Appl* **8,** 284–295.

Wang, Y. (2017). Spatial distribution of high copy number plasmids in bacteria. *Plasmid* **91**, 2-8.

Wang, Y., Penkul, P. and Milstein, J. N. (2016). Quantitative Localization Microscopy Reveals a Novel Organization of a High-Copy Number Plasmid*. Biophys J* **111**, 467–479.

Wang, J., Stephan, R., Karczmarczyk, M., Yan, Q., Hachler, H. and Fanning, S. (2013). Molecular characterization of bla ESBL-harbouring conjugative plasmids identified in multi-drug resistant *Escherichia coli* isolated from food-producing animals and healthy humans. *Front Microbiol* **4**, 188.

Wang, D., Yang, J., Fang, M., He, W., Zhang, Y., Liu, C. and Zhou, D. (2017). Characterization of the Novel In1059 Harbouring VIM Gene Cassette. *Antimicrob Resist Infect Control* **6**, 50.

Warnes, A. and Stephenson, J. R. (1986). The insertion of large pieces of foreign genetic material reduces the stability of bacterial plasmids. *Plasmid* **16**, 116-123.

Watanabe, T. (1969). Bacterial episomes and plasmids. *CIBA Found Symp* 81-97.

Waters, V. L.(1999). Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance. *Front Biosci* **4**, D433–56.

Wein, T., Hülter, N. F., Mizrahi, I. and Dagan, T. (2019). Emergence of Plasmid Stability Under Non-Selective Conditions Maintains Antibiotic Resistance. *Nat Commun* **10**, 2595.

Weissman, S. J., Chattopadhyay, S., Aprikian, P., Obata-Yasuoka, M., Yarova-Yarovaya, Y., Stapleton, A., Ba-Thein, W., Dykhuizen, D., Johnson, J. R. and Sokurenko, E. V. (2006). Clonal Analysis Reveals High Rate of Structural Mutations in Fimbrial Adhesins of Extraintestinal Pathogenic *Escherichia coli*. *Microbiol* **59**, 975–988.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C., Ochman, H. and Achtman, M. (2006). Sex and Virulence in *Escherichia coli*: An Evolutionary Perspective. *Mol Microbiol* **60**, 1136-1151.

World Health Organization (2020). Global antimicrobial resistance and use surveillance System (GLASS) Report - Early Implementation 2020. Geneva, Switzerland.

Xu, Z., Li, L., Shi, L., Shirtliff, M. E. (2011). Class 1 Integron in Staphylococci. *Mol Biol Rep* **38**, 5261–79.

Yarmolinsky, M. B. (1995). Programmed Cell Death in Bacterial Populations. *Science* **267**, 836–837.

Zhang, C., Xu, X., Pu, S., Huang, S., Sun, J., Yang, S. and Zhang, L. (2014). Characterization of Carbapenemases, Extended Spectrum β-Lactamases, Quinolone Resistance and Aminoglycoside Resistance Determinants in

carbapenem-Non- Susceptible *Escherichia coli* from a Teaching Hospital in Chongqing, Southwest China. *Infect Genet Evol* **27**, 271–276.

Chapter 2 Genomic diversity of strains isolated from patients living in the same geographical region.

2.1 Abstract

The frequency of multi-drug resistant human pathogenic bacteria has significantly increased all over the globe. The three lineages of the phylogroup B2 have been equally implicated in the cause of these extra-intestinal infections, but their antimicrobial profiles differ. Because resistance genes are often borne on plasmids, investigation of the differences in the genomic content of the plasmids found in the lineages of the B2 phylogroup, and how the plasmids promote the spread of resistance genes was carried out using various bioinformatics tools. Number of strains analysed were 80 (ST131), 76 (ST73) and 79 (ST95). This study shows that there were key differences in the plasmids content within these lineages. The number of mobilisable plasmids per strain varied among lineages (Wilcoxon Test: $X^2 = 8.73$, $P > 0.0127$). ST131 strains had an average fewer mobilisable plasmids than ST73 or ST95 strains. By contrast ST73 strains were less likely to host a conjugative plasmid compared to strains of the other STs (Wilcoxon Test: $X^2 = 50.28$, $P < 0.0001$). No difference among STs in the total number of plasmids harboured per strain was detected (Median Test: $X^2 = 1.819$, $P > 0.4027$). The total plasmid DNA content per strain varied significantly (ANOVA: $F = 14.49$, $P > 0.0001$), ST131, n=80, ST73, n= 76 and ST95, n=79) among the lineages. The average plasmid DNA content of ST131 strains was 147kbp, 164kpb for ST95 strains and 112kbp for ST73. Among strain differences within an ST also varied among STs (Brown-Forsythe Test: $F = 5.05$, $P > 0.007$). ST73 strains exhibited the greatest among strains variation and ST95 strains the least, a result that reflects

the fact that many ST73 strains lacked conjugative plasmids. IncF plasmids that play an important role in conjugal transfers were very common in all the *E. coli* strains analysed especially IncFII and IncFIB plasmids, but this class of plasmids was more frequently found in the strains of the ST131 lineage. There were diversities in the resistance determinants in ST131 lineage based on the presence of different classes of antimicrobial resistance genes as well as the frequent occurrence of fluoroquinolone resistance determinant set this lineage apart from the other two (STs 73 and 95) lineages. Most fluoroquinolone resistant isolates were most adapted to O25H4:ST131:*fimH30*. O16H5:*fimH41* is another host characteristics common in ST131 strains. ST73 had no fluoroquinolone positives in any of its isolates. The most prevalent serotype observed in the strains of the ST95 lineage was 01:H7 with *fimH2*7 while the most prevalent serotype in ST73 lineage was 06:H1 with *fim30.* Integrons were found associated with resistance genes within the strains of ST131 lineage, but they were few when compared to the other lineages. Most of the ST131 strains associated with multiple resistance genes and integrons were detected in the F-plasmid sequence type F29:A-:B10. ST95 and ST73 lineages share a major F-plasmid sequence type: F51:A-:B10 and integrons were observed to dominate this F-plasmid sequence type, but fewer multiple antimicrobial genes were noted. F51:A-:B10 sequence type was not found in any of the strains of the ST131 lineage. Plasmid encoded virulence genes and bacteriocins were common in these lineages. Most of these genes are involved in DNA transfer, cytolysis, and defense responses. Overall, distinct genetic diversity was observed in the plasmids of the B2 lineages. Thus, the unique differences within the plasmids of these lineages may be able to account for their differing antimicrobial profiles.

2.2 Introduction

Escherichia coli clonal groups ST131, ST95 and ST73 frequently account for a large proportion of extra-intestinal *E. coli* infections, occurring at equal rates in both clinical and community settings in many parts of the world (Nicolas-Chanoine *et al*., 2008; Riley, 2014). However, one of the lineages, ST131, has been implicated in several studies as the major cause of multidrug resistance in extra-intestinal *E. coli* infections (Kallonen *et al.*, 2017; Johnson *et al*., 2016; Skov and Monnet, 2016; Nicolas-Chanoine *et al*., 2014; Price *et al*., 2013; Coque *et al*., 2008; Nicolas-Chanoine *et al*., 2008). Limited information exists on the reasons why the two other lineages, ST95 and ST73, are largely drug susceptible (Stephens *et al*., 2017; Kallonen *et al*., 2017; Gordon *et al*., 2017; Bengtsson *et al*., 2012). Specifically, how these drug susceptible strains evolve to cause pandemics, and whether there are other biological characteristics that are unrelated to drug resistance.

Plasmids are double-stranded DNA molecules, which are naturally found in bacteria, archaea, and recently, in some eukaryotes. They are mobile genetic elements that exist independently of the chromosome, and are self-replicating (Martinez-freijo *et al*., 1998). They are known to encode genes required by the cell, especially during adverse conditions. Virulence genes, antibiotic resistance genes, genes conferring resistance to toxic heavy metals, and many others enable the cell to exploit specific environmental situations (Thomas and Summers, 2008). Resistant plasmids are known to encode multiple antimicrobial resistance genes (Levin, 1995). Since plasmids are important determinants of antimicrobial resistance in *E. coli,* understanding the differences in the genomic diversity of the three lineages with emphasis on their plasmid content, frequency of occurrence, and other intrinsic factors within their genome, becomes

essential to provide clues as to why STs 73 and 95 are drug susceptible despite that all three of them are globally pathogenic.

There has been a significant increase in the use of bacterial whole-genome sequences, due to readily available next generation DNA sequencing techniques, especially with regards to its use in monitoring the outbreak of resistant pathogens. Various software tools were utilised to analyse draft genomes of the strains of phylogroup B2 lineages from clinical samples obtained from Canberra Hospital, Australia over a period of time. These techniques examine the genetic characteristics of the genomic make-up of each lineage in order to understand their differences.

The aim of this project was to investigate the differences in the genomic diversity of the plasmids in the phylogroup B2 lineages using various bioinformatics tools.

2.3 Materials and Methods

2.3.1 Bioinformatics Tools

Plasmid communities of 235 draft whole genome sequenced (WGS) data of *E. coli* isolates from blood, urine and faeces taken from people living in the Canberra region of Australia were analysed These strains served as a proxy for strains that had spread globally (Merino *et al*., 2016; Gordon, *et al*., 2015; Blyton *et al*., 2014; Gordon, 2010; Yumuk *et al*., 2008; Gordon, *et al*., 2005; Gordon, *et al*., 2002). There were 76 isolates of ST73, 80 of ST131, and 79 of ST95. Three recently developed bioinformatics tools designed to detect contigs associated with plasmid DNA were used to determine the

number of plasmids within the strain of each lineage. The bioinformatics tools used were Mlplasmid (Sergio Arredondo-Alonso *et al*., 2018), Plascope (Royer *et al*., 2018), and MobSuite (Robertson and Nash, 2018).

Mlplasmid is a binary classifier used to predict short-read contigs either as a plasmid or chromosome derived. It is a species - specific classifier that employs pentamer frequencies. The binary classifier assigns each contig to either plasmid or chromosome using a probability threshold. Plascope on the other hand, is based on a centrifuge software tool that uses a metagenomic classifier and a custom database. It assigns sequences based on matches against the database, and then sorts them into classes. MobSuite is an application tool for identifying, grouping, reconstructing, and typing of plasmids from whole genome assemblies. It uses a reference database method to identify contigs of plasmid origin and then cluster them into classes based on an internal grouping strategy. MobSuite accepts fasta formatted genome assemblies.

The outputs predictions from the three bioinformatics tools were combined to enhance accuracy and wherever there is a consensus among all the tools for plasmid, such contig is accepted as plasmid derived. The outputs from these analyses were also incorporated into pipeline such as BLAST (basic local alignment search tool) to further confirm the results. This was done by uploading the assembled genomes in fasta format to the BLAST programme. BLAST examines for regions of similarity among biological sequences. The programme compares nucleotide or protein sequence, to sequence databases, and then, calculates the statistical significance (National Center for Biotechnology Information [NCBI]).

Incompatibility (Inc) groups were identified by uploading the assembled genomes in fasta-format to the centre for genomic epidemiology (CGE) (www.genomicepidemiology.org). The presence of plasmids in a strain is based on finding matches to known Inc groups using the application tool, PasmidFinder 2.1 (Carattoli *et al.,* 2014). The PlasmidFinder web tool is able to identify plasmids in total or partial sequenced isolates of bacteria. The presence of antibiotic resistance genes was determined by uploading assembled genomes in fasta format to ResFinder 3.1 (Zankari *et al*., 2012). Serotyping was done by using the SerotypeFinder 2.0 tool (Joensen *et al*., 2015). The *fimH* type was identified by uploading the genomes to FimTyper version 1.0 (Roer *et al*., 2017). ResFinder, VirulenceFinder, SerotypeFinder and the FimTyper are all tools available on the CGE website.

The draft sequenced data were also analysed using other bioinformatics tools such as MacSyFinder, CONJScan and IntegronFinder, MacSyFinder is an application tool use to model and detect macromolecular systems, and genetic pathways in protein datasets. In prokaryotes, these systems have evolutionary conserved properties: they are made up of conserved components and are encoded in compact loci (conserved genetic architecture). The user models these systems with MacSyFinder to reflect the conserved features, and to allow their efficient detection. Detection of a system is based on the presence of a given amount of mandatory and accessory components in the pre-defined genetic architecture. CONJScan allows for the detection of conjugative systems in genomes. A conjugative system is defined by the presence of a relaxase, a Type IV coupling protein and a Type IV secretion system (T_4SS). There are eight different types of T_4SS. They are defined by the phylogeny of the only ubiquitous protein, Virb4 found in T_4SS. Conjugation system Types T, G, F, and I for single cell

bacteria such as *E. coli* were detected using this application tool. IntegronFinder was used to detect the presence of integrons in the bacterial genomes.

2.3.2 Statistical analysis of genomic data

The number of plasmids, plasmid DNA content, F- plasmid sequence types, *fimH* type, serotype, number of incompatibility groups, resistance genes associated with the ExPEC groups, presence of integrons, virulence factors, and type of conjugation system were determined and compared using one-way analysis of variance (ANOVA) and Median test at 95% probability. The significant differences were noted. All analyses and graphic designs were done in JMP 15, and Microsoft excel software.

2.4 Results

2.4.1 Serotype and *fimH* type diversity

Varying distributions of serotypes and *fimH* types within the different lineages (fig. 2.1) (pearson X^2 test: P< 0.0001) were observed. Serotypes O25:H4 and O16:H5 were the only serotypes detected in ST131 lineage. Serotypes 01:H7, 050 /02:H7, and 050 / 02:H4 were most prevalent in ST95, while 06:H1 was the most prevalent serotype in ST73 lineage. Some ST95 strains exhibited the serotype 025:H4.

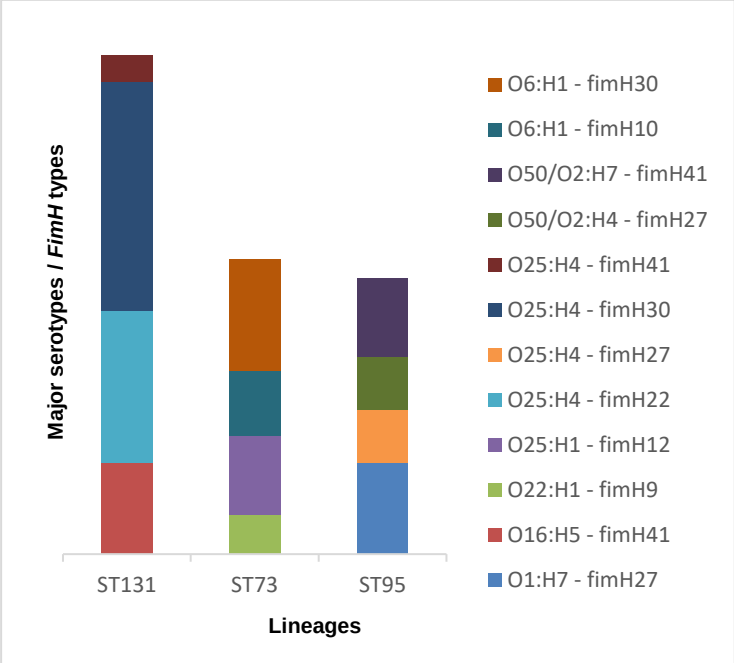

Figure 2.1 Distribution of serotypes / *fimH* types by lineages

Six *fimH* types were identified among the strains of the ST131 lineage, although, majority of ST131 strains belonged to *fimH30* with 025:H4 serotype. Ten different *fimH* types were identified in the strains of ST95 lineage, but *fimH27* and *fimH41* with O1:H7 and O50/02:H7 serotypes respectively were more prevalent. ST73 lineage harboured about fifteen different *fimH* types. *FimH30* with *O6:H1* serotype was most prevalent, this was followed closely by *fimH12* and *fimH10* with 025:H1 and *O6:H1* serotypes, respectively.

2.4.2 Number of plasmids per strain

MobSuite attempts to infer the number of plasmids present in each bacterial isolate using draft assemblies, and if the plasmids are conjugative or mobilisable. The results of the MobSuite analysis revealed that the number of mobilisable plasmids per strain varied among lineages (Wilcoxon Test: $X^2 = 8.73$, $P > 0.0127$) (fig. 2.3). ST131 strains

had on average fewer mobilisable plasmids than ST73 or ST95 strains. By contrast ST73 strains were less likely to host a conjugative plasmid compared to strains of the other STs (Wilcoxon Test: $X^2 = 50.28$, $P < 0.0001$) (fig. 2.2). No difference among STs in the total number of plasmids harboured per strain was detected (Median Test: $X^2 = 1.819$, $P > 0.4027$) (fig. 2.4), with 2 as the median number of plasmids per strain. While each strain of ST131 harbour more than one plasmid, about 5% of ST73 and 2% of ST95 strains harbour only one plasmid (fig. 2.4).

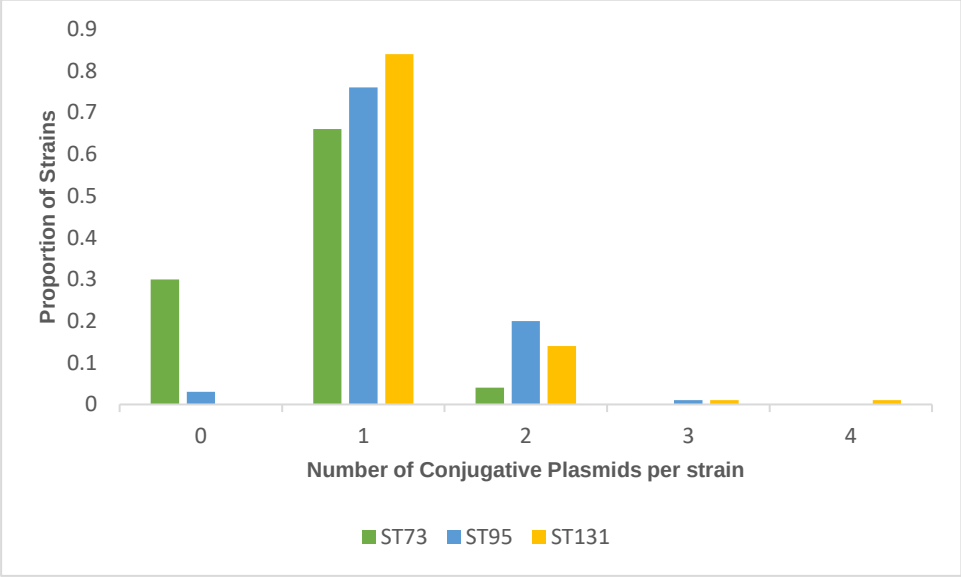

Figure 2.2 **Number of conjugative plasmids per strain as inferred by the MobSuite tool with respect to ST membership.**

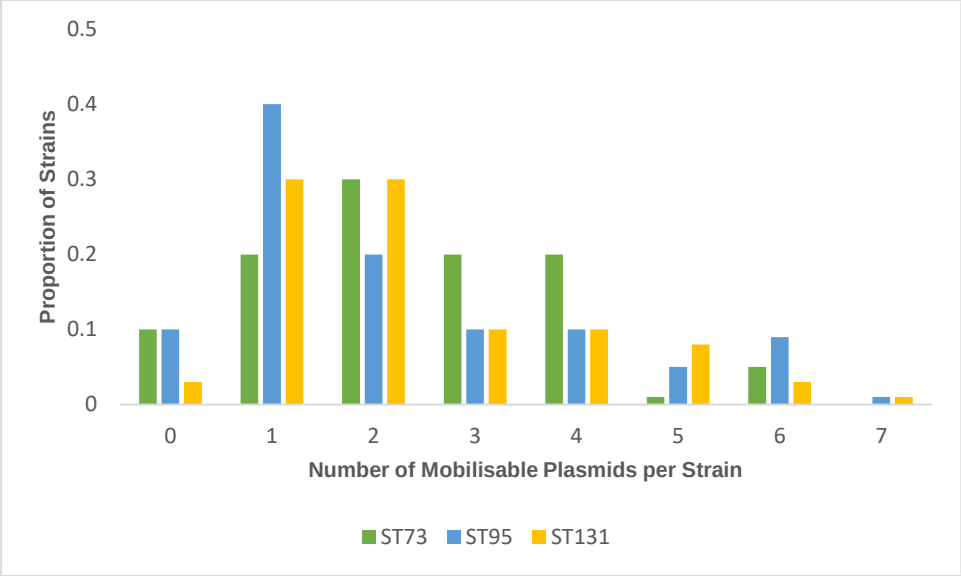

Figure 2.3 **Number of mobilisable plasmids per strain as inferred by the MobSuite tool with respect to ST membership.**

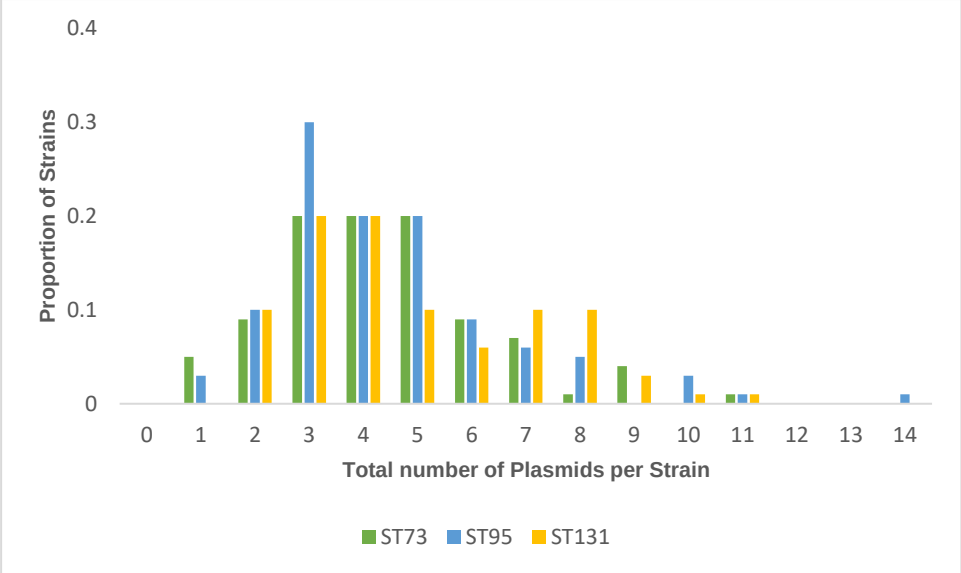

Figure 2.4 **Number of plasmids per strain as inferred by the MobSuite tool with respect to ST membership**

2.4.3 Total plasmid DNA content per strain

The total plasmid DNA content of each strain was inferred by integrating the results of the tools MobSuite, Mlplasmid and Plascope. Total plasmid DNA content per strain

varied with ST (ANOVA: F = 14.49, P < 0.0001). The average plasmid DNA content of ST73 strains was 112kbp, 164kpb for ST95 strains and 147kbp for ST131 (figs. 2.5, 2.6 and 2.7). As well as the STs differing in their total plasmid DNA content, the among strain variation within an ST also varied among STs (Brown-Forsythe Test: F = 5.05, P > 0.007). ST73 strains exhibited the greatest among strains variation and ST95 strains the least, a result that reflects the fact that many ST73 strains lacked conjugative plasmids.

Total plasmid DNA per strain increased with the number of plasmids present in a strain, but the magnitude of the relationship varied among the STs (ANOVA: ST, F = 10.71, P < 0.0001; Number of Plasmids, F = 63.21, P < 0.0001; ST * Number of Plasmids, F = 6.46, P > 0.0019) (fig 2.8).

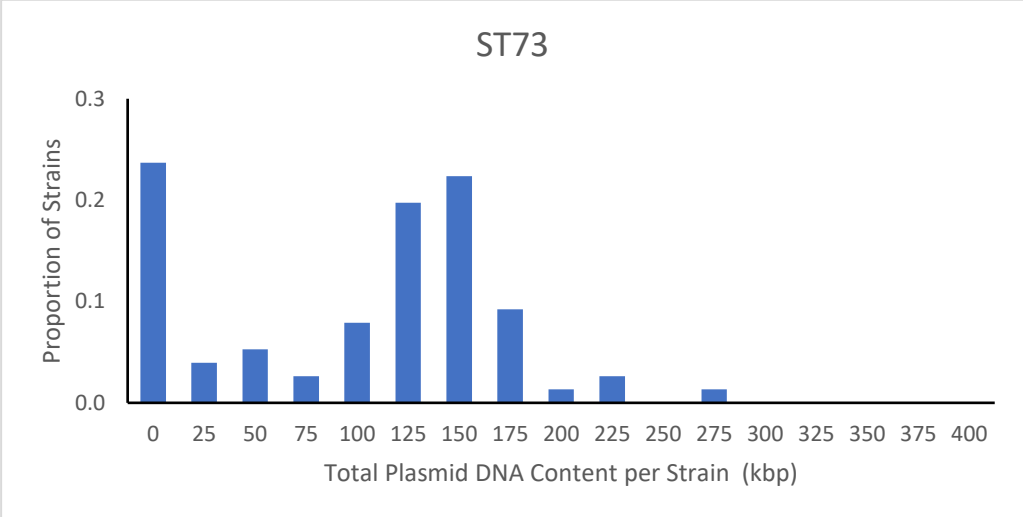

Figure 2.5 Total plasmid DNA content per strain with respect to ST membership

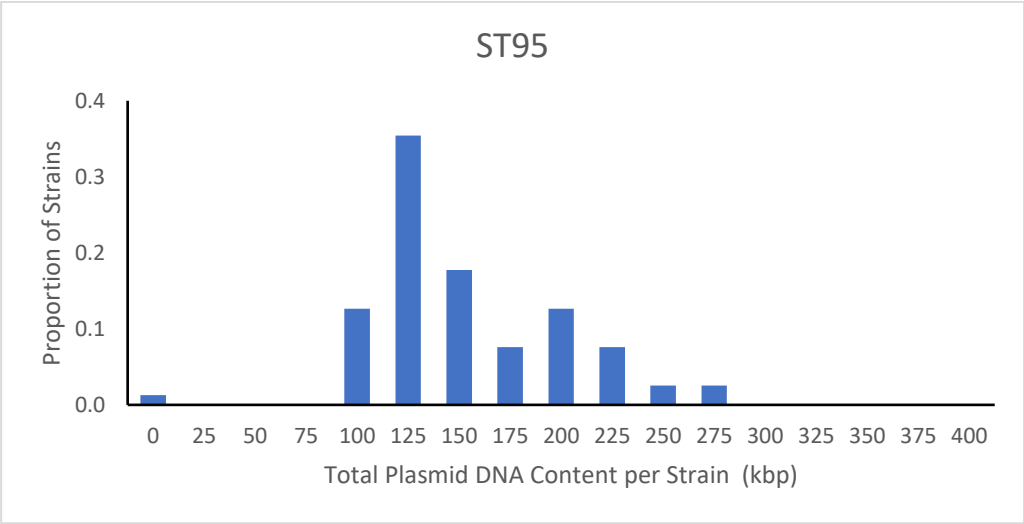

Figure 2.6 **Total plasmid DNA content per strain with respect to ST membership**

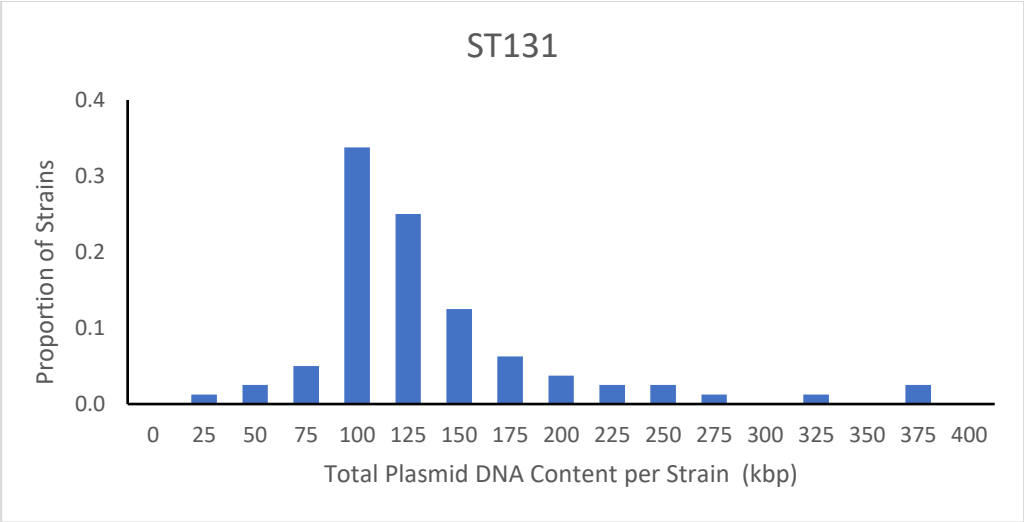

Figure 2.7 **Total plasmid DNA content per strain with respect to ST membership**

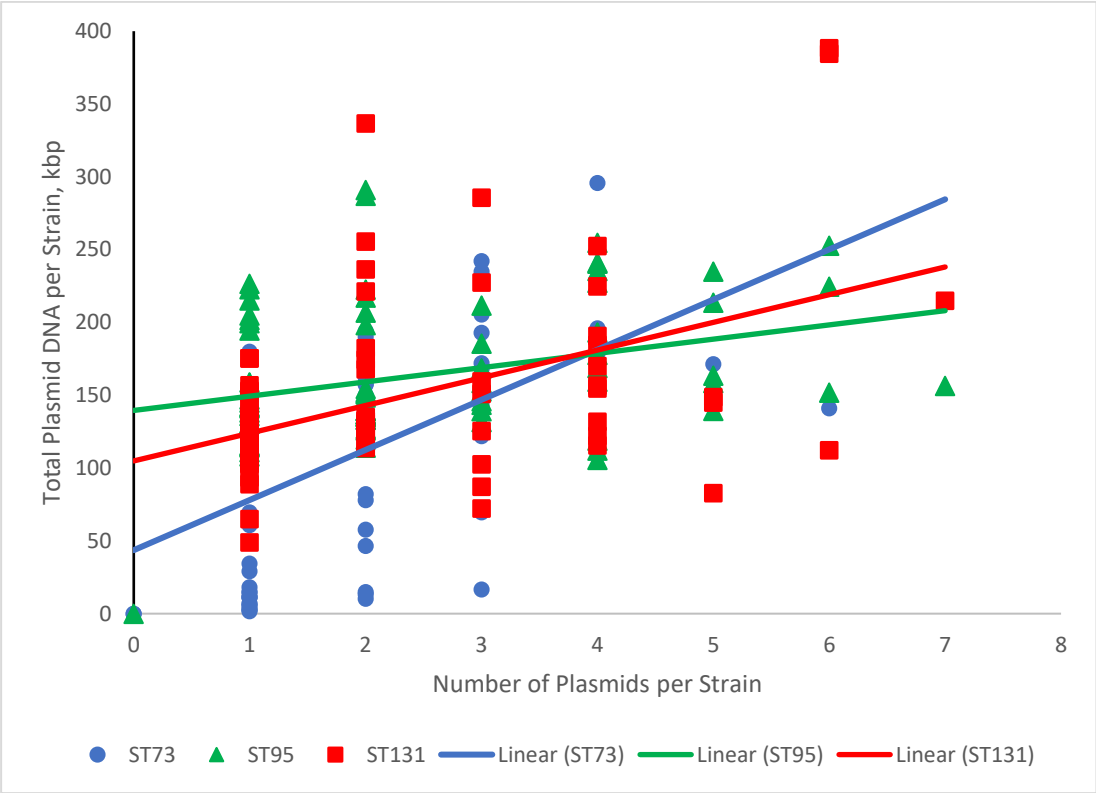

Figure 2.8 **The relationship between the total number of plasmids inferred per strains and the total amount of plasmid DNA inferred per strain with respect to the ST membership of the strains**

2.4.4 Incompatibility groups diversity

Thirteen incompatibility groups were detected in total. The incompatibility groups Inc12, Inc13, IncH, IncN, IncX1, and IncX4, and InCY were uncommon, and none were detected in more than six strains. The average number of Inc groups varied with ST (Wilcoxon Test: $X^2 = 29.12$, $P < 0.0001$), with an average of three Inc groups detected in ST73 strains compared with an average of four in STs 95 and 131.

The common Inc groups were non-randomly distributed with respect to ST (Contingency Table: $X^2 = 53.13$, $P < 0.0001$) (fig. 2.9). IncFIA was over- represented in ST131 strains, while InC1 and IncQ1 were over-represented in ST95 strains.

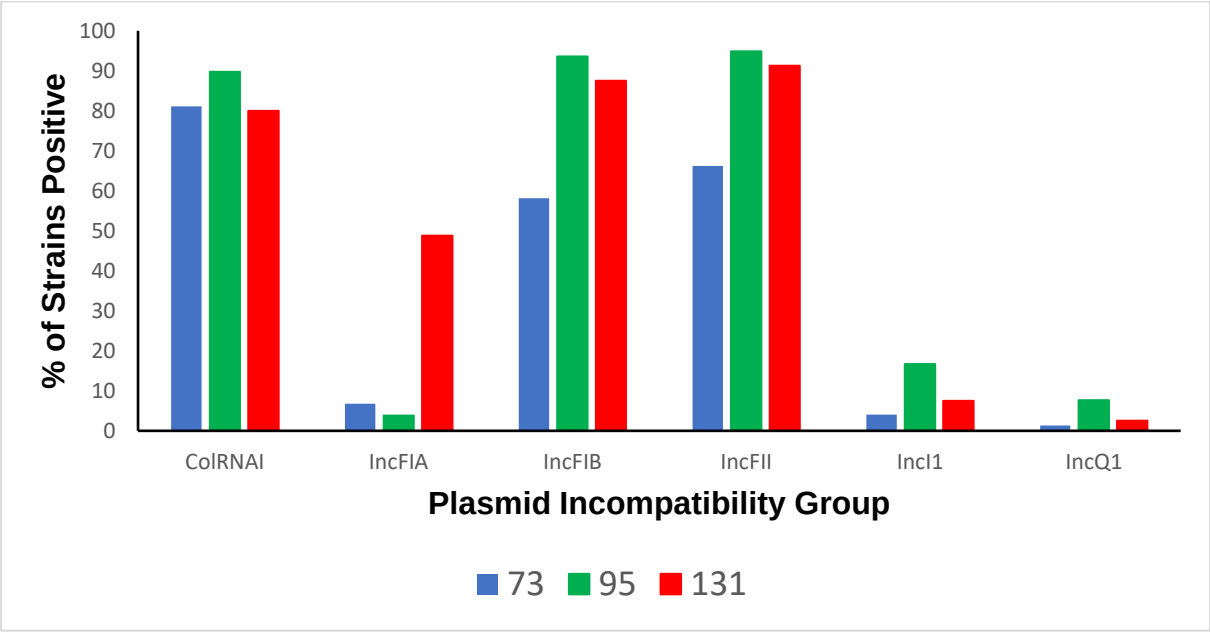

Figure 2.9 **Distribution of the most common incompatibility groups with respect to sequence type**

2.4.5 Distribution of conjugation systems

The probability of detecting one or more of the conjugation systems F, G, I and T varied among the STs (Contingency Table: X^2 = 38.01, P < 0.001). These conjugation systems were not detected in 55% of ST131 strains, and were absent in 43% of ST73 strains, while only 13% of ST95 strains lacked a conjugation system. All the analysed plasmids of the *E. coli* strains were observed to carry diverse conjugation systems (fig. 2.10) (Pearson X^2 test: P< 0.0001), Type G, F, T and I. The conjugation Type T system was more prevalent in ST131, n=80 strains. However, very few of the strains of ST95, n=79 and ST73, n=76 carried this conjugation type. The conjugation Type G system was not found in either of the strains of ST131 or ST73 lineages. However, this conjugation Type G system was observed in some strains of ST95 lineage. Conjugation Type F was very common in the strains of the three lineages: ST95 has

the highest distribution, followed by ST73 and then, ST131. Overall, all the ExPEC strains analysed harboured at least, two conjugation systems (fig. 2.10).

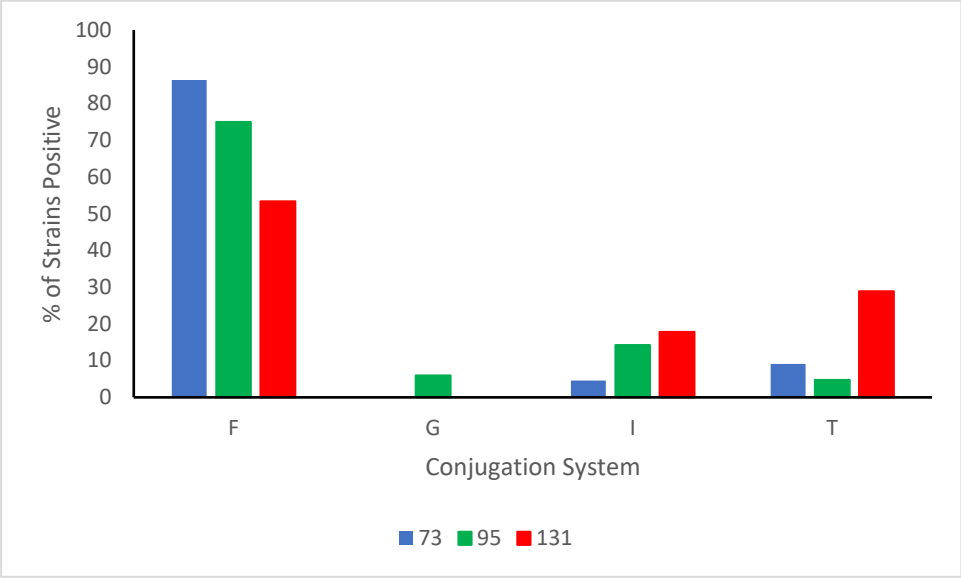

Figure 2.10 Distribution of conjugation systems with respect to sequence type

2.4.6 Plasmid encoded virulence profiles of the ExPEC lineages

Varying number of plasmid encoded virulence genes and bacteriocins were detected in the ST lineages (contingency table 2.1 and 2.2). The number of virulence factors detected per strain varied by ST ($P < 0.05$) (single factor Anova). Plasmid-encoded traits consistently present in the strains of all the lineages includes enterotoxin tieB protein (*senB),* enterobacterial complement resistance protein (*traT*), aerobactin synthase (*iucC*), protein iss (*iss*), tellurium resistance protein (*terC*), aerobactin receptor (iutA), outer membrane receptor protein (*ompT*) and SitA protein (*sitA*). Cytotoxin necrotizing factor 1 (*cnF1*), common in patients with urinary tract infections was prevalent in the strains of ST73 than the other STs (contingency table 2.1).

In contrast to ST131 strains, plasmid encoded colicin E1 (*cea*) was very common in ST73 and ST95 strains. Microcin B17 (*mcbA*) were detected in some strains of ST95, none of these were found in the other STs (contingency table 2.2). This result also shows that the presence of virulence genes and bacteriocins are not strongly correlated with the presence of antibiotic resistance genes (contingency table 2.2).

Virulence genes	**ST131**	**ST73**	**ST95**
cnf1	5%	53%	9%
iha	70%	75%	3%
iroN	5%	79%	33%
iss	49%	99%	91%
sat	41%	71%	0%
senB	64%	53%	41%
ireA	0%	30%	52%
sitA	73%	90%	87%
iutA	19%	22%	23%
traT	86%	61%	81%
ompT	99%	97%	87%
terC	98%	100%	79%
iucC	34%	17%	22%
afaD	9%	30%	22%
hlyF	6%	0%	18%
etsC	4%	0%	37%

Table 2-1 Virulence determinants per strain

Bacteriocins	ST131	ST73	ST95
cea	1%	28%	35%
cvaC	1%	1%	9%
cia	5%	0%	10%
cib	1%	0%	0%
mcbA	5%	0%	0%

Table 2-2 Correlation heat map showing plasmid encoded bacteriocins per strain

2.4.7 Antibiotic resistance determinants associated with the plasmids of the ExPEC lineages.

No difference was found in the number of antimicrobial resistance genes on plasmids among the lineages (Median Test: X^2 =0.52, P > 0.77), ST131, n=80, ST73, n= 76 and ST95, n=79 (fig. 2.11). However, varying classes of multi-drug resistance genes were detected in ST131 in contrast to the other lineages (contingency table 2.3). Fluoroquinolone genes were only detected in the strains of ST131, none of these genes were found in the strains of ST73 and ST95. In addition, trimethoprim resistance gene, (*dfrA17*), sulphonamide resistance gene (*sul1*), tetracycline resistance gene, (*tetA*) and aminoglycoside modifying genes (*aac(3)-IId* and *aadA5*) were over-represented in ST131 than in the other lineages. Aminoglycoside modifying genes, *aadA2* and *aada1*, sulphonamide resistance gene, *sul2*, tetracycline resistance gene, *tetB* and trimethoprim resistance genes, *dfra12* and *dfra5* were common in ST73 and ST95. However, the β-lactamase genes, *blaTEM-1A, blaTEM-1B* and trimethoprim resistance genes, *dfrA1* and *dfrA3* were prevalent in the strains of ST73. Although, the presence of genes encoding TEM, and SHV were observed in ST73 strains, there was no blaCTX-M positives in any of the isolates.

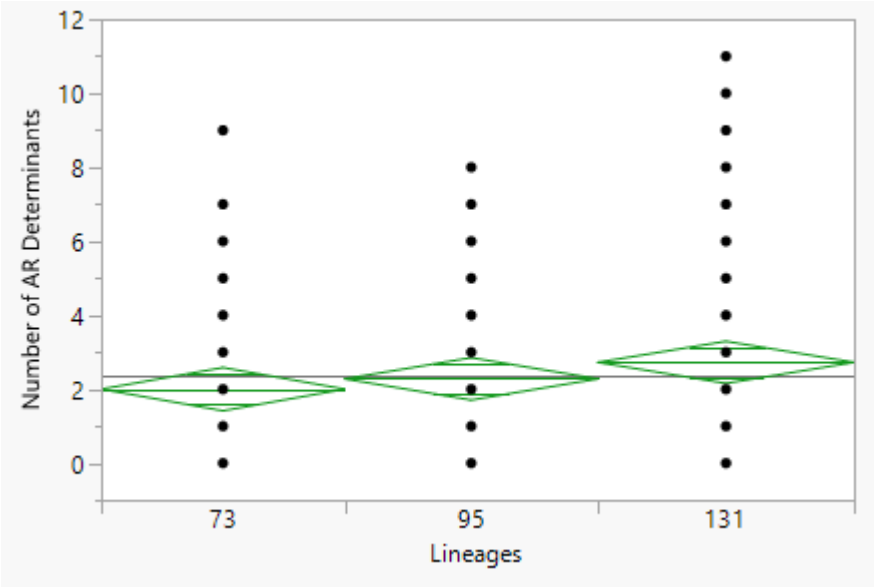

Figure 2.11 Antimicrobial resistance genes by ST lineages

Antibiotics class	Antibiotic Resistance Gene	ST131 n=80	ST73 n=76	ST95 n=79	P>p
Aminoglycosides	aph(3')-Ia	18%	9%	8%	0.11117
	aph(6)-Id	18%	9%	6%	0.06632
	aac(3)-Iid	18%	1%	3%	0.00005
	aadA1	4%	29%	18%	0.00009
	aadA2	1%	1%	6%	0.09979
	aadA5	19%	1%	5%	0.00016
	ant(2'')-Ia	1%	0%	0%	0.3811858
Beta-lactams	blaTEM-1A	0%	11%	0%	0.00014
	blaTEM-1B	38%	59%	35%	0.00413
	blaTEM-1C	4%	3%	13%	0.01871
	blaTEM-30	0%	0%	1%	0.3741656
	blaSHV-2	0%	1%	0%	0.3528275
	blaCTX-M 15	21%	0%	3%	2.00E-07
Fluoroquinolones	aac(6')-Ib-cr	13%	0%	0%	0.00002
chloramphenicol	catB4	10%	0%	0%	0.00027
	catA1	6%	3%	0%	0.06677
	cmlA1	1%	0%	5%	0.07394
Sulphonamides	sul1	29%	28%	23%	0.66719
	sul2	19%	15%	20%	0.62658
	sul3	0%	0%	4%	0.0497622
Tetracycline	tetA	20%	3%	6%	0.00049
	tetB	8%	13%	22%	0.03777
	tetC	0%	1%	1%	0.5971279

Antibiotics class	Antibiotic Resistance Gene	ST131 n=80	ST73 n=76	ST95 n=79	P>p
Trimethoprim	dfra12	1%	1%	6%	0.09979
	dfrA14	3%	0%	1%	0.38364
	dfrA17	20%	1%	8%	0.00027
	dfrA1	0%	7%	0%	0.00452
	dfrA5	1%	1%	8%	0.04156
	dfra7	3%	1%	0%	0.37614
Total Number of Resistance Genes		218	152	159	0.11
Multi-drug resistance (%)		57%	36%	43%	0.0000013

Table 2-3 Major classes of antibiotic resistance genes within the lineages

2.4.8 Comparison of ST131 serotypes with respite to antibiotic resistance profiles and fimHtypes.

The two common serotypes found in ST131 are associated with specific *fimH* types and numerous antibiotic resistance genes. Resistance genes with O25H4:*fimH30* is over- represented in this lineage. O16H5:*fimH41* is also common. O25H4:*fimH22 and* O25H4:*fimH41* are associated with less resistance genes. No strain of ST131/*fimH30* belong to O16:H5, rather, all strains with O16:H5 belong only to *fimH41*.

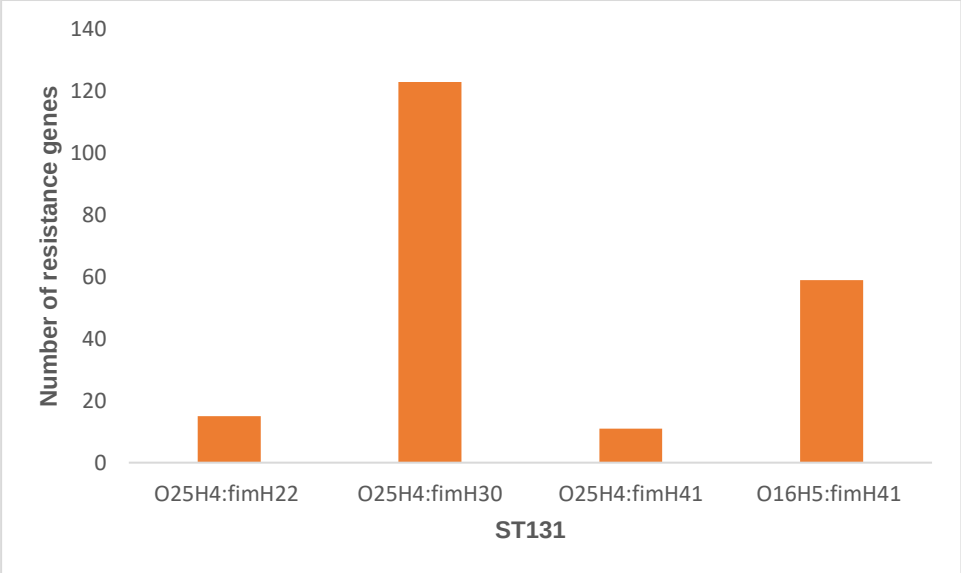

Figure 2.12 Distribution of resistance genes and *fimH* types in the serotypes of ST131 lineage

2.4.9 F- Plasmid sequence type diversity in the ExPEC lineages.

Forty-nine F-plasmid sequence types were detected in total. Distribution of F-plasmid sequence types were varied among the lineages (Pearson X^2 test: P< 0.0001). Although, the diversity was more prevalent in ST131, associated with about 23 different F- plasmid sequence types. The most common sequence types were F29:A-:B10, F1:A2:B20, F31:A4:B1 and F2:A1:B1. ST95 lineage has about sixteen different F-plasmid sequence types; F51:A-:B10, F24: A-:B1, F29:A-:B10 and F2:A:B1 were prevalent in most strains. Ten different F- plasmid sequence types were found in ST73. the three commonly encountered in the lineage were the F51:A-:B10, F-:A-:B- and F29: A-:B10 were very common (fig. 2.13). F-:A-:B- was always found in strains where no plasmids were detected and so, was excluded from the analysis.

Although, F31:A4:B1 was observed in few strains of ST95, F1:A2:B20 was not found in any strain of ST73 and ST95. Likewise, even though F29:A-:B10 was common in the three lineages, F51:A-:B10 was represented only in the strains of ST73 andST95 (fig. 2.13).

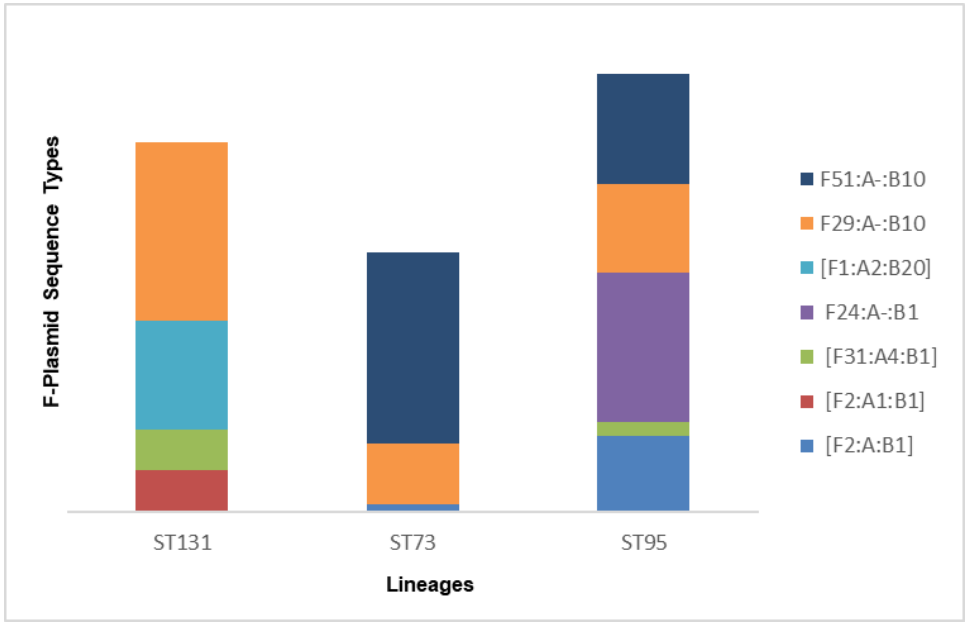

Figure 2.13 Major F-plasmid sequence types by lineages

2.4.10 Distribution of integron positive strains within the ExPEC lineages

Inferred MobSuite result shows no variation in the frequency of integron carriage among the three lineages (Median Test: X^2 =0.57, P > 0.75) (fig. 2.14). However, differences were detected in the frequency of integron carriage within the F- plasmid sequence types (Pearson X^2 test= 82.69, P< 0.0001). Strains of ST131 with F29:A:B10, were found to harbour 10% of the total integrons while strains of ST73 and ST95 with F-plasmid sequence type, F51:A:B10 have 64% of the total integrons (fig. 2.15). These integrons were identified to be associated with numerous resistance genes in these STs with *BlaTEM-1B, Sul1* and aadA1 being very prevalent (fig. 2.16). However, resistance genes were more represented with the integrons of ST131 in contrast to the other STs (fig. 2.17).

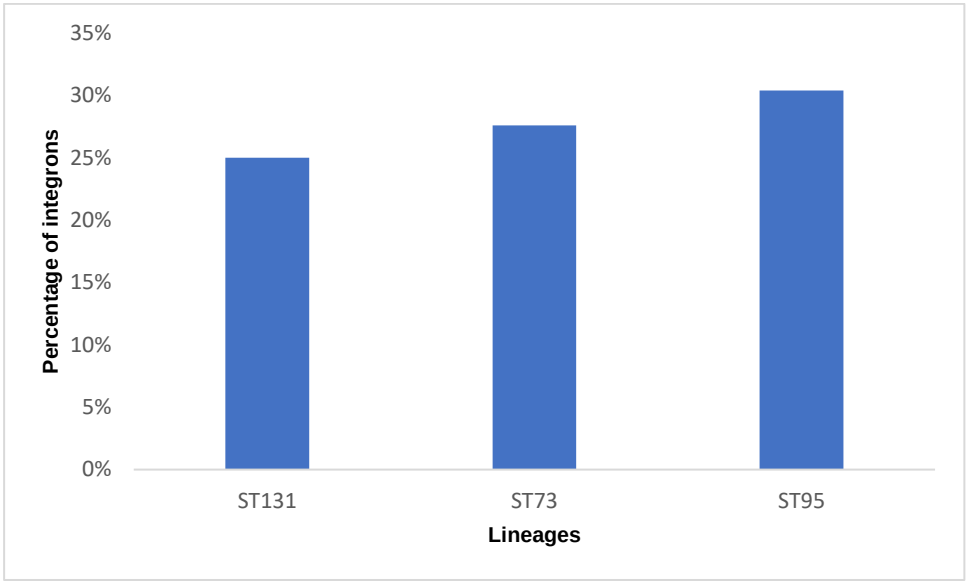

Figure 2.14 Integron Carriage among the lineages

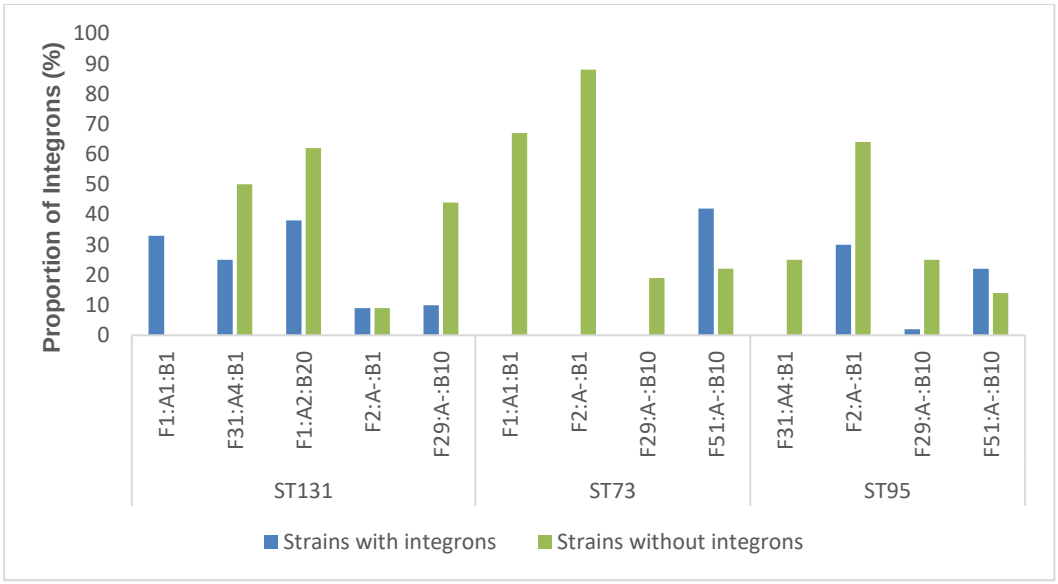

Figure 2.15 Common F-plasmid sequence types associated with integrons in the lineages

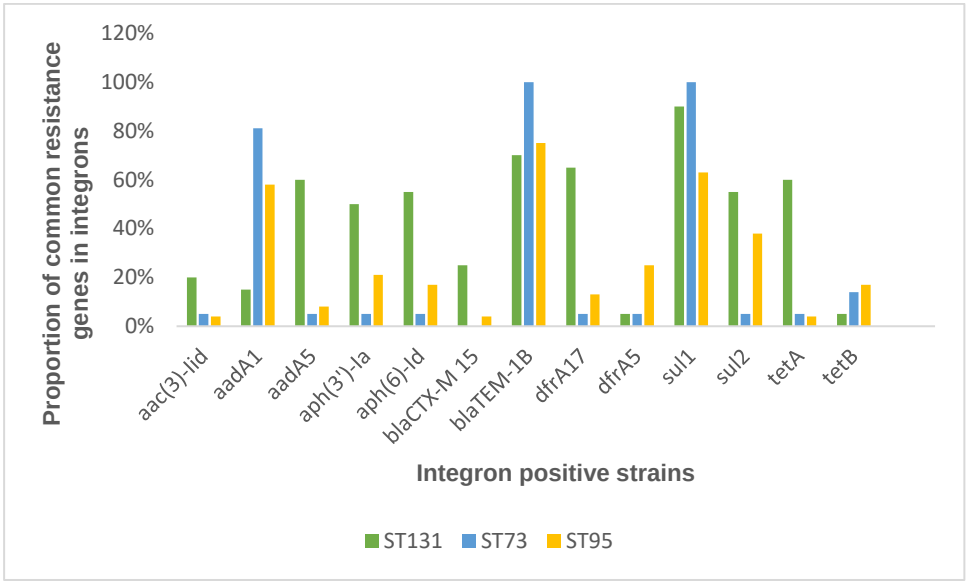

Figure 2.16 Resistance genes associated with integrons in the ST lineages

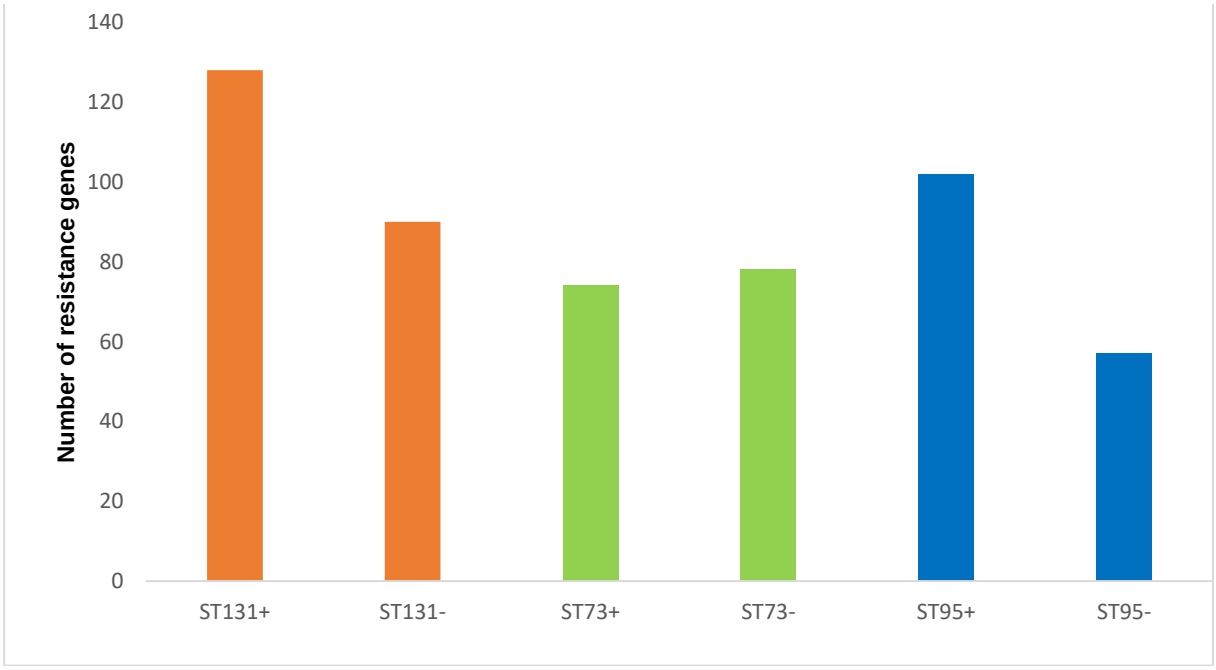

Figure 2.17 Distribution of resistance genes with / without integrons among the lineages

2.5 Discussion

In this study, draft whole genome sequence data of 235 *E. coli* isolates from various clinical samples were subjected to different bioinformatics tools to detect differences in the genomic diversity of the plasmids within the strains of the three lineages of phylogroup B2: ST131, ST95 and ST73, that cause wide variety of intestinal and extra-intestinal infections.

The inferred results from Mobsuite application tool showed that the three lineages are genetically diverse and carried different arrays of plasmids with distinct features. Plasmids are one of the important mobile genetic elements that convey antimicrobial resistance in *E. coli.* They ensure the movement of genetic material between bacteria and often mediate resistance to multiple antimicrobials. (Stokes and Gillings, 2011). Plasmids enhance the adaptation and fitness of a bacterial host to unfavourable environments (Touchon *et al*., 2020). Although, there was no definite difference in the carriage of plasmids between the different STs in this study, each of the investigated *E*. *coli* strains harboured at least one plasmid, and in most cases, two or more smaller plasmids were found in the same strain. This result correlates with the report that demonstrated that all *E*. *coli* strains carried at least one plasmid, and more than five co-resident plasmids in some instances (Sherley *et al*., 2003). Small mobilisable plasmid DNAs found in most strains of ST73 as well as in some strains of ST95 lineage in this study is not surprising, because other reports have shown that mobilisable plasmids are relatively small, often less than 10kb (Garcillán-Barcia *et al*., 2009). Most of these mobilisable plasmids were found to have plasmid DNA that are less than 40kb. Although such plasmids cannot transfer genetic material, their co-existence, often with conjugative plasmids, enables their transfer during bacterial conjugation (Riemmann

and Haas, 1993). There is thus the possibility that co-integration of these low weight mobilisable plasmids were responsible for the results in the present study. In contrast to mobilisable plasmids, conjugative plasmids prevalent in ST131 and ST95 strains are often large, with low copy numbers and are common in the strains of *E*. *coli* (Smillie *et al*., 2010). Conjugative plasmids are essential for conjugal transfer of genes responsible for bacterial adaptation and evolution (Schröder and Lanka, 2005; Lawley, *et al*., 2003). Fertility plasmids, IncFII and IncFIB were over-represented in the lineages. These replicons play very essential functions in bacterial conjugation. They are known to encode different determinants that ensure host fitness and the plasmid replication and maintenance (Van Melderen and Saavedra, 2009). This combination of replicons has been previously reported as a multi-replicon on a single plasmid (Pierre-Emmanuel *et al*., 2020; Villa *et al*., 2010; Johnson *et al*., 2007), probably due to co-integration of two different plasmids. Conjugation systems are also an important component for bacterial conjugation. Although majority of the strains in the three lineages carried the conjugation system Type F, only about 55% of the strains of ST73 have conjugation system. This should be expected because 29% of ST73 plasmid DNAs do not belong to any IncF group, necessary for conjugal transfer to occur and many low weight plasmid DNAs were also identified in this lineage. This study further demonstrated that only about 45% of the strains in ST131 harbour conjugation systems, even when many numerous conjugative plasmids and IncF replicons were detected. This, therefore, suggests that conjugation may not be the only means by which genetic material is being transferred in the strains of ST131.

Multiple classes of antimicrobial resistance genes and the presence of fluoroquinolone determinant found only in the plasmids of the strains of ST131 lineage especially those

that harboured conjugative plasmids distinguished this lineage from the other two lineages. The high prevalence of O25:H4:*fimH30:*ST131 clonal group in this study was not new because similar findings have been reported (Giedraitiene *et al*., 2017; Kallonen *et al*., 2017; Dias *et al*., 2009). These studies reported that most strains belonging to *E. coli* ST131 lineage are of serotype O25:H4 and are highly antibiotic resistant. The strains encoding such serotype are also known to harbour multiple antibiotic resistance gene *ESBL CTX-M-15* (Lau *et al.,* 2008). These strains of *E. coli* O25:H4:ST131:*CTX-M*-15 have also been previously detected in other countries (Merino *et al*., 2016; Yumuk *et al*., 2008), and are known to be associated with an increased capacity for plasmid uptake with high plasmid diversity. In addition, blaCTX−M gene was found associated with fluoroquinolone resistance genes, and this was common in most strains of ST131 (Tsukamoto *et al*., 2013). Aminoglycoside resistance detected in multiple proportions in the strains of ST131 especially in *ESBL*-producing Enterobacteriaceae is of high public health interest because misuse of this antibiotic could promote the spread of *ESBL* genes (Cantón and Ruiz-Garbajosa, 2011).

Non-random distribution of plasmid encoded virulence genes and bacteriocins were identified in the strains of these lineages. Complement resistance protein involved in regulation of conjugation as well as many adhesin genes were over-represented in the lineages. Cytotoxin necrotizing factor 1 (*cnF1*), common in patients with urinary tract infections (Island *et al*., 1998; Falbo *et al*., 1992), was detected in abundance in each strain of ST73 than the others. Plasmid encoded *colicin E1* and *colicin V* were also common in the lineages. Although *colicin ia*, *ib* and *microcin B17* were only found in

STs 131 and 95, all the bacteriocins are involved in cytolysis and defense response. These results further suggest the pathogenicity of each of these lineages.

Integrons were identified in F-plasmid sequence types, F31:A4-:B1 and F1:A2-:B20, but these were only detected in the strains of ST131. F-plasmid sequence type is a classification type for closely related fertility plasmids based on the type of genes present in them. Although, integrons were found in some strains of ST131 and ST95 with F29:A-:B10 and F2:A-:B1, they were uncommon. Integrons were however well represented in all strains with F51:A-:B10. Integrons are genetic elements that can capture and express antimicrobial resistance genes through gene cassettes encoded in them. Various studies have reported the presence of integrons in plasmids (Partridge *et al*., 2009; Rowe-Magnus *et al*., 2003). Integrons promote the evolution of plasmids by their roles in conjugal gene transfers (Mazel, 2006). *Beta-lactams*, *blaTEM-1b*, sulphonamides, *sul1* and the aminoglycosides *aadA1* genes were often associated with both F51:A:B10 and F29:A:B10. β-lactamases mediate resistance to β-lactam antimicrobial agents in *E. coli*. It is primarily responsible for ampicillin resistance. It inactivates the antibiotic by hydrolysing the β- lactam ring (Livermore, 1995). Sulphonamides have a broad range of antimicrobial activity against Gram-negative organisms. Sulphonamides are widely used to treat urinary tract infections, but resistance to the antibiotic has increased over the years with many Integron borne *Sul1* on plasmids being reported (Bean, 2005; Enne *et al*., 2001). These results indicate that there is a relationship between certain antimicrobial resistance genes, F-plasmid sequence type and integrons. All strains harbouring an integron were found to be resistant to at least one antibiotic. This correlation may also reflect resistance cassette arrays within the integrons. These results agree with previous studies from

clinical isolates (White *et al*., 2001; Martinez-Freijo *et al*., 1998; Jones *et al.*, 1997), that detected numerous antimicrobial genes in associated with sequence type F29:A:B10. This sequence type was also found associated with the *fimH* types *fimH41*, *fimH22* and *fimH30* known with drug resistance genes. Even though most of the strains of the ST131 lineage with sequence type F29:A:B10 do not possess integrons, the few that do harbour more of the *dfrA*17 and *aadA*5 genes. These gene cassettes are known to both confer resistance to trimethoprim and encode an aminoglycoside-3Q-adenylyl-transferase (Xia *et al*., 2016). The greater abundance of these gene cassettes in ST131 highlights the capacity of this lineage to evolve multi-drug resistance mechanisms that could continue to have negative impacts in the future use of therapeutic drugs (White *et al*., 2000). Furthermore, genomes with small plasmid DNAs lack integrons as observed in the strains of ST73 lineage, while genomes with large conjugative plasmid DNAs accumulated most of the integrons in the ST131 and ST95 lineages. This further demonstrates the capacity of large conjugative plasmids to encode mobile genetic elements as opposed to mobilisable plasmids and thus, the variations in the antimicrobial profiles of the three lineages. Several factors such as virulence determinants, presence of adhesins, resistance determinants can influence the evolution of plasmids. These likewise can impact the unequal distribution of plasmid incompatibility groups (Sherley *et al*., 2004) as observed in this study. STs 131 and 95 seem to have some identical resistance profiles, but our study showed that they have diverse genetic backgrounds. The multi-drug resistance classes, coupled with fluoroquinolone resistance genes in ST131 strains, likely enhanced the capacity of this lineage to be highly multi-drug resistant and globally pathogenic.

Overall, in relation to F- plasmid sequence type, number of resistance determinants varied. Presence of more mobilisable plasmids encoding *bla*TEM-1 and Aada1 genes, associated with many integrons in F-plasmid sequence type F51:A:B10, found only in ST95 and ST73 despite their antibiotic susceptibility, may be connected to their continued persistence, pathogenicity, pandemicity. But with respect to the lineage of the bacterial host (STs), differences in the genomic diversity of the plasmids revealed that there was no significant difference in the proportion of antimicrobial resistance genes present within these STs, however, there was diversity in the antimicrobial profiles within the strains of ST131 because of the presence of different classes of antimicrobial resistance genes and fluoroquinolone resistance genes. However, factors other than F-plasmid sequence type superiority may still be in action. Therefore, an extensive understanding of ST73 and ST95 compared to ST131 would be further exploited experimentally.

2.6 Conclusion

Distinct genetic diversity was observed in the plasmids of the phylogroup B2 lineages. Conjugative plasmids are essential for conjugal transfer of genes responsible for bacterial adaptation and evolution. ST73 is less likely to carry conjugative plasmids because of the numerous small mobilisable plasmid DNAs found in most strains of this lineage as well as in some strains of ST95. Genomes with small plasmid DNAs lack integrons as observed in the strains of ST73 lineage, while genomes with large conjugative plasmid DNAs accumulated most of the integrons in the ST131 and ST95 lineages. This further demonstrates the capacity of large conjugative plasmids to encode mobile genetic elements as opposed to mobilisable plasmids and thus, the

variations in the antimicrobial profiles of the three lineages. Although, fertility plasmids, IncFII and IncFIB were over-represented in the lineages and are very important in bacterial conjugation, more than one third of ST73 strains are devoid of these replicons. Variations were not observed in the number of resistance genes, plasmid encoded virulence genes and bacteriocins in the three lineages. But multiple classes of antimicrobial resistance genes and the presence of fluoroquinolone determinant were only found in the conjugative plasmids of the strains of ST131. These likely enhanced the capacity of this lineage to be highly multi-drug resistant and globally pathogenic. The key F-plasmid sequence type, F51:A-:B10, found only in ST73 and ST95 strains encode many integrons. While integrons were less common in most strains of ST131 and some of ST95 with F29:A-:B10 and F2:A-:B1, most ST131 encoding integrons were found to be associated with many antibiotic resistance determinants. Although, some strains of ST73 and ST95 may lack resistance genes because of the absence of plasmids, most strains encode beta-lactams, sulphonamides and the aminoglycosides genes often associated with both F51:A:B10 and F29:A:B10. These results therefore show a relationship between specific F-plasmid sequence types, antimicrobial resistance genes, and integrons. Therefore, the unique differences within the plasmids of these lineages may be able to account for their differing antimicrobial profiles.

2.7 References

Bean, D. C., Livermore, D. M., Papa, I. and Hall, L. M. C. (2005). Resistance Among *Escherichia coli* to sulphonamides and other antimicrobials now little used in man. *J Antimicrob Chemo* **56**, 962–964.

Bengtsson, S., Naseer, U., Sundsfjord, A., Kahlmeter, G. and Sundqvist, M. (2012). Sequence Types and Plasmid Carriage of Uropathogenic *Escherichia coli* Devoid of Phenotypically Detectable Resistance. *J Antimicrob Chemother* **67**, 69–73.

Blyton, M. D., Banks, S. C., Peakall, R., Lindenmayer, D. B. and Gordon, D. M. (2014). Not All types of host contacts are equal when it comes to *E. coli* transmission. *Ecol Lett* **17**, 970–978.

Cantón, R. and Ruiz-Garbajosa, P. (2011). Co-resistance: An Opportunity for the Bacteria And Resistance Genes. *Curr Opin Pharmacol* **11**, 477–485.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Aarestrup, F. M. and Hasman, H. (2014). *In Silico* Detection and Typing of Plasmids Using Plasmidfinder and Plasmid Multilocus Sequence Typing. *Antimicrob Agents Chemother* **58**, 3895–3903.

Coque, T. M., Novais, A., Carattoli, A., Poirel, L., Pitout, J., Peixe, L., Baquero, F. and Canton, R., Nordmann, P. (2008). Dissemination of Clonally Related *Escherichia coli* Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15. *Emerg Infect Dis* **14**, 195–200.

Dias, R. C. S., Marangoni, D. V. Smith, S. P. Alves, E. M. Pellegrino, F. L. P. C. Riley, L. W. and Moreira, B. M. (2009). Clonal Composition of *Escherichia coli* Causing Community-acquired Urinary Tract Infections in the State of Rio De Janeiro, *Brazil. Microb Drug Resist* **15,** 303-308*. Microbiol Spectr* **2**, 04.

Enne, V. I., Livermore, D. M., Stephens, P. and Hall, L. M. C. (2001). Persistence of sulphonamide resistance in *Escherichia coli* in the UK despite national prescribing restriction. *Lancet* **357**, 1325-1328.

Falbo, V., Famiglietti, M. and Caprioli, A. (1992). Gene block encoding production of cytotoxic necrotizing factor 1 and hemolysin in *Escherichia coli* isolates from extraintestinal infections. *Infect Immun* **60**, 2182-2187.

Garcillán-Barcia, M. P., Francia, M. V. and de la Cruz, F. (2009). The Diversity of Conjugative Relaxases and its application in Plasmid Classification. *FEMS Microbiol Rev* **33**, 657–687.

Giedraitiene, A., Vitkauskiene, A., Pavilonis, A., Patamsyte, V., Genel, N., Decre, D. and Arlet, G. (2017). Prevalence of O25b-ST131 Clone among *Escherichia coli* Strains Producing CTX-M-15, CTX-M-14 and CTX-M-92 beta-lactamases. *Infect Dis* **49**,106–112.

Gordon, D. M. (2010). Strain typing and the ecological structure of *Escherichia coli*. *J. AOAC International* **93**, 974–984.

Gordon, D. M., Bauer, S. and James R. Johnson, J. R. (2002). The genetic structure of *Escherichia coli* populations in primary and secondary habitats*. Microbiology*. **148**, 1513–1522.

Gordon, D. M., O'Brien, C. L. and Pavli, P. (2015). *Escherichia coli* Diversity in the Lower Intestinal Tract of Humans. *Environ. Microbio. Reports*. **7**, 642–648.

Gordon, D. M., Geyik, S., Clermont, O., O'Brien, C. L., Huang, S., Abayasekara, C., Rajesh, A., Kennedy, K., Collignon, P., Pavli, P., Rodriguez, C., Johnston, B. D., Johnson, J. R., Decousser, J. W. and Denamur, E. (2017). Fine-Scale Structure Analysis Shows Epidemic Patterns of Clonal Complex 95, a Cosmopolitan Escherichia coli Lineage Responsible for Extraintestinal Infection. *mSphere*. pp 2.

Gordon, D. M., Stern, S. E. and Colington, P. J. (2005). Influence of the Age and Sex of Human Hosts on the Distribution of *Escherichia coli* ECOR Groups and Virulence Traits. *Microbiology* **151**, 15–23.

Island, M. D., Cui, X. L. and Foxman, B. (1998). Cytotoxicity of hemolytic, cytotoxic necrotizing factor 1-positive and -negative *Escherichia coli* to human T24 bladder cells Infect Immun 66, 3384-3389.

Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M. and Scheutz, F. (2015). Rapid And Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* **53**, 2410-26.

Johnson, T. J., Danzeisen, J. L., Youmans, B., Case, K., Llop, K., Munoz-Aguayo, J., Flores-Figueroa, C., Aziz, M., Stoesser, N., Sokurenko, E., Price, L. B. and Johnson, J. R. (2016). Separate F-Type Plasmids Have Shaped the Evolution of the *H*30 Subclone of *Escherichia coli* Sequence Type 131. *Clinical Science and Epidemiology mSphere,* **1** (4).

Johnson, T. J., Wannemuehler, Y. M., Johnson, S. J., Logue, C. M., White, D. G., Doetkott, C., and Nolan, L. K. (2007). Plasmid Replicon Typing of Commensal and Pathogenic *Escherichia coli* isolates. *Applied and environmental microbiology* **73**, 1976-1983.

Jones, M., Peters, E., Weersink, A. M., Fluit, A. and Verhoef**,** J**. (**1997). Widespread Occurrence of Integrons Causing Multiple Antibiotic Resistance in Bacteria. *Lancet* **349,** 1742-1743.

Kallonen, T., Brodrick, H. J, Harris, S. R, Coriander, 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**, 1437-1449.

Lau, S. H., Kaufmann, M. E., Livermore, D. M., Woodford, N., Willshaw, G. A., Cheasty, T., Stamper, K., Reddy, S., Cheesbrough, J., Bolton, F. J., Fox, A. J. and Upton, M. (2008). UK Epidemic *Escherichia coli* Strains A-E, with CTX-M-15 Beta-Lactamase, all belong to the International O25:H4-ST131 Clone. *J Antimicrob Chemother* **62**, 1241-1244.

Lawley, T. D., Klimke, W. A., Gubbins, M. J. and Frost, L. S. (2003). F- Factor Conjugation is a True Type IV Secretion System. *FEMS Microbiol Lett* **224**, 1–15.

Levin, B. R. (1995). Conditions for the Evolution of Multiple Antibiotic Resistance Plasmids: a Theoretical and Experimental Excursion. *Population Genetics of Bacteria*, Pp.175–192. Edited by S. Baumberg. Cambridge/New York: Cambridge University Press.

Livermore, D. M. (1995). Lactamases in laboratory and clinical resistance. *Clin Microbiol Rev* **8**, 557–584.

Martinez-Freijo, P., Fluit, A. C., Schmitz, F. J., Grek, V. S., Verhoef, J. and Jones, M. E. (1998). Class I Integrons in Gram-Negative Isolates from Different European Hospitals and Association with Decreased Susceptibility to Multiple Antibiotic Compounds. J. *Antimicrob Chemother* **42,** 689-696.

Mazel, D. (2006). "Integrons: Agents of Bacterial Evolution". *Nature Reviews Microbiology.* **4,** 608–620.

Merino, I., Shaw, E., Horcajada, J. P., Cercenado, E., Mirelis, B., Pallarés, M. A., Gómez, j., Xercavins, M., Martínez-Martínez, L., De Cueto, M., Cantón, R., Ruiz-Garbajosa, P., Bunshow, E., Sánchez-Carrillo, C., Padilla, B., Benito, N., Gamallo, R., Riera, M., Calbo, E., Fariñas, M. C., Gonzalo, M., Pascual, A., Rodríguez-Baño, J. and Pintado, V. (2016). CTX-M-15-H30Rx-ST131 Subclone is one of the main causes of Healthcare-Associated ESBL-Producing *Escherichia coli* Bacteraemia of Urinary origin in Spain. *J Antimicrob Chemother* **71**, 2125–2130.

Nicolas-Chanoine, M. H., Bertrand, X. and Madec, J. Y. (2014). *Escherichia coli* ST131, an Intriguing Clonal Group. *Clin Microbiol Rev* **27**, 543–574.

Nicolas-Chanoine, M. H., Blanco, J., Leflon-Guibout, V., Demarty, R., Alonso, M. P., Caniça, M. M., Park, Y. Y., Lavigne, J. P., Pitout, J. and Johnson, J. R. (2008). Intercontinental Emergence of *Escherichia coli* Clone O25:H4-ST131 Producing CTX-M-15. *J Antimicrob Chemother* **61**, 273-281.

Partridge, S. R., Tsafnat, G., Coiera, E. and Iredell, J. R. (2009). Gene Cassettes and Cassette *Arrays in Mobile Resistance Integrons. FEMS Microbiol Rev **33**,* 757–784*.*

Pierre-Emmanuel, D., Ludovic, M., Nicolas, R., Arnaud, F. and Michel-Yves, M. (2020). Analysis of COMPASS, a New Comprehensive Plasmid Database

Revealed Prevalence of Multi-replicon and Extensive Diversity of IncF Plasmids. *Frontiers in Microbiology* **11,** 483.

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-B-Lactamase-Producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30. Rx. *MBio* **4**, 377-13.

Riemmann, C. and Haas, D. (1993). Mobilisation of chromosomes and Non-Conjugative Plasmids by Co-integrative Mechanisms in Bacterial Conjugation, pp. 137–188. Edited by D. B. Clewell. New York: Plenum.

Riley, L.W. (2014). Pandemic Lineages of Extraintestinal Pathogenic *Escherichia coli*. *Clin Microbiol. Infect* **20**,380–390*.*

Robertson, J. and Nash, J. (2018). MOB-suite: Software Tools for Clustering, Reconstruction and Typing of Plasmids from Draft Assemblies. *Microbial genomics **4*** (8).

Roer, L., Hansen, F., Stegger, M., Sönksen, U. W., Hasman, H. and Hammerum, A. M. (2017). Novel mcr-3 Variant, Encoding Mobile Clistin Resistance, in an ST131 *Escherichia coli* Isolate from Bloodstream Infection, Denmark. 2014. Euro Surveill 3, 22-31.

Rowe-Magnus, D. A., Guerout, A. M., Biskri, L., Bouige, P. and Mazel, D. (2003). Comparative Analysis of Superintegrons: Engineering Extensive Genetic Diversity in the Vibrionaceae. *Genome Res **13***, 428–442.

Royer, G., Decousser, J. W., Branger, C., Dubois, M., Médigue, C., Denamur, E., and Vallenet, D. (2018). PlaScope: A Targeted Approach to Assess the Plasmidome from Genome Assemblies at the Species Level. *Microbial genomics*, **4**(9).

Schröder, G. and Lanka, E. (2005). The Mating Pair Formation System of Conjugative Plasmids-A Versatile Secretion Machinery for Transfer of Proteins and DNA. *Plasmid* **54**, 1–25.

Sergio Arredondo-Alonso, Malbert, R. C., Johanna, C. B., Tess, D. V., Janetta, T., Jukka, C., Rob, J. L., and Anita C. S. (2018). Mlplasmids: A User-Friendly Tool to Predict Plasmid and Chromosome Derived Sequences for Single Species. *Microbial Genomics* **4.**

Sherley, M., Gordon, D. M. and Collignon, P. J. (2003). Species Differences in Plasmid Carriage in the Enterobacteriaceae. Short Communication*.* *ScienceDirect* **49**, 79-85.

Sherley, M., Gordon, D. M. and Collignon, P. J. (2004). Evolution of multi-resistance plasmids in Australian clinical isolates of Escherichia coli. *Microbol* **150**, 1539-46.

Skov, R. L. and Monnet, D. L. (2016). Plasmid-Mediated Colistin Resistance (mcr-1 gene): Three Months Later, the Story Unfolds. *Euro Surveill* **21**, 30155.

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. and de la Cruz, *F.* (2010). "Mobility of Plasmids. *Microbiology and Molecular Biology Reviews. 74,* 434–452.

Stephens, C. M., Adams-Sapper, S. Manraj Sekhon, Johnson, J. R. and Riley, L. W. (2017). Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. *mSphere*, **2** (2). DOI: 10.1128/mSphere.00390-16.

Stokes, H. W. and Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *Fems Microbiology Reviews* 35**,** 790-819.

Thomas, C. M. and Summers, D. (2008). Bacterial Plasmids. In eLS, (Ed.). https://doi.org/10.1002/9780470015902.a0000468.pub2.

Touchon, M., Perrin, A., de Sousa, J. A. M., Vangchhia, B., Burn, S., o'Brien, C. L., Denamur, E., Gordon, D. and Rocha, E. P. (2020). Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *Biorxiv* 02.12.945709.

Tsukamoto, N., Ohkoshi, Y., Okubo, T., Sato, T., Kuwahara, O., Fujii, N., Tamura, Y. and Yokota, S. (2013). High Prevalence of Cross-Resistance to Aminoglycosides In Fluoroquinolone-Resistant *Escherichia coli* Clinical Isolates. *Chemother* **59**, 379– 384.

Van Melderen, L. and Saavedra De Bast, M. (2009). Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities? *PLoS Genet* **5**, e1000437.

Villa, L., García-Fernández, A., Fortini, D. and Carattoli, A. (2010). Replicon Sequence Typing of IncF Plasmids Carrying Virulence and Resistance Determinants. *Journal of Antimicrobial Chemotherapy* **65**, 2518–2529.

White, P. A., McIver, C. J., Deng, Y. M. *and* Rawlinson, W. D. *(*2000*).* Characterisation of Two New Gene Cassettes, aadA5 and dfrA17. *FEMS Microbiol Lett **182**,* 265–269.

White, P. A., McIver, C. J. and Rawlinson, W. D. (2001). Integrons and Gene Cassettes in the *Enterobacteriaceae*. *Antimicrob Agents Chemother* **45,** 2658-2661.

Xia, W., Xu, T., Qin, T., Li, P., Liu, Y., Kang, H., Gu, B. and Ma, p. (2016). Characterization Of Integrons and Novel Cassette Arrays in Bacteria from Clinical Isolates In China, 2000-2014. *J Biomed Res* **30**, 292–303.

Yumuk, Z., Afacan, G., Nicolas-Chanoine, M. H., Sotto, A. and Lavigne, J. P. (2008). Turkey: A Further Country Concerned by Community-Acquired *Escherichia coli* Clone. O25-ST131 Producing CTX-M-15. *J Antimicrob Chemother* **62**, 284–288.

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. *J Antimicrob Chemother* **67**, 2640–2644.

Chapter 3 Plasmid - Host Dynamics

3.1 Abstract

Plasmids are essential for bacteria evolution and adaptation. They promote bacterial adaptation by spreading important ecological characteristics among lineages using the mechanism of conjugation. Stability of these plasmids within the host is known to depend on the benefits of the plasmids, and its cost to the bacterial host. To understand the rate at which these plasmids are being transferred, how these plasmids are maintained within *E. coli* lineages, STs 131, 73 and 95 and the impacts of the plasmid carriage on the host, there is the need to investigate them under laboratory conditions using various analytical tools. In these analyses, the rate in which plasmids are transferred using plasmid transfer experiments were performed. To examine the impacts of the plasmids on transfer rates, genome wide association study was carried out using Prokka, Roary and Random Forest application tools. We then experimentally determined the frequency at which the plasmids are lost within the transconjugants strains using plasmid loss measurements tests via serial dilution techniques. These transconjugants were derived from earlier conjugal transfers of the donor *E. coli* (ancestral) strains of ST131, ST73 and ST95 lineages to K-12 recipient cell. We further examined the effects of the plasmid carriage on the bacterial host using same transconjugants strains, by calculating the relative fitness of plasmid carrying and plasmid-free strains using competition experiments. The conjugal transfers performed revealed that all 93 strains analysed were able to transfer their plasmids by conjugation, although rates of transfer differed. No diversity was observed in the rate of plasmid transfers among the ST lineages (Median Test: X^2 =0.19, P = 0.90),

however small differences were observed in the rate of plasmid transfer in relation to the F-plasmid sequence types. This showed that the rate of plasmid transfer did not depend on host genetic background. The random forest approach however revealed that there was a direct relationship between rate of plasmid transfers and the F-plasmid sequence types based on encoded variable genes (*traC*, *traF*, *traH*, *traR*) (Median Test: $P < X^2=0.0001$), found during the analysis. These *tra* genes are always found in the *tra* operon. The distribution of the *tra* genes among the ST lineages also revealed that F29:A-:B10 and F51:A-:B10 plasmids harbour the highest proportion of all *tra* genes recorded. Although, plasmids were still transferred in the absence of the multiple *tra* genes, the presence of the *tra* genes positively enhanced the rates of plasmid transfer. Plasmid stability results showed that plasmid dynamics differ among the three lineages. After seven days transfers, significant differences in the plasmid stability between the lineages were observed (Wilcoxon test: X2= 25.06, P<0.0001). Large differences were also observed in plasmid stability in relation to F-plasmid sequence types. (Wilcoxon test: X2= 16.71, P<0.0001). The effects of the plasmids on host fitness showed a fairly small but significant difference between the bacterial host fitness and the plasmid diversities (Kruskal-Wallis Tests: $X^2 = 8.67$, $P< 0.013$), but there was much significant difference between the F-plasmid sequence types diversities and the host fitness (Kruskal-Wallis Tests: $X^2 = 7.63$, $P< 0.005$). The results shown here suggest, that the genetic background of the host cell and the F-plasmid sequence types associated with the plasmids, influenced the conjugal transfer, plasmid stability, host fitness and the overall adaptation and evolution of the plasmids of the *E. coli* lineages.

3.2 Introduction

While the movement of plasmids from one lineage to another is horizontal gene transfer, HGT is generally thought of, in terms of changes in the genes on the bacterial chromosome as a result of HGT. The transfer of plasmids through bacterial conjugation has increasingly advanced the dissemination of antibiotic resistance genes within *E. coli* strains (Uma *et al*., 2009). The rates at which plasmids are transferred depends on the initial transfer initiated during conjugation in donor cells, and its successful integration into the recipient cells (Thomas and Nielsen, 2005).

Bacterial evolution, adaptation and dissemination frequently relies on the acquisition of new auxiliary genes encoded on mobile genetic elements (Frost *et al*. 2005), such as plasmids. Such conjugative plasmids are essential because of their capacity to disseminate useful ecological attributes in bacterial lineages (Hall *et al*., 2016; Jain *et al*., 2003; Ochman *et al*., 2000).

The dynamics and stability of conjugative plasmids in bacterial communities are essential in explaining bacterial evolution (Thomas and Nielsen, 2005). The fate of a plasmid in a bacterial host population can be influenced by numerous determinants, such as cost of plasmid carriage on bacterial host cell, rate of plasmid conjugation and segregation, and the robustness of selection of auxiliary genes harboured on plasmids, which usually vary depending on the environment (Slater *et al*. 2008; Bergstrom *et al*., 2000; Simonsen 1991). Plasmid instability can lead to increase in the number of cells that are devoid of plasmid compared to the plasmid carrying cells, resulting in plasmid loss (Harrison *et al*., 2015; San Millan *et al*., 2014; De Gelder *et al*., 2007). However, some plasmids can persistently exist in bacterial host over a long period. To fully

understand how this happens is difficult, particularly when the plasmids exist in the absence of plasmid encoded determinants. This is pertinent because maintenance of plasmids by bacterial host is often burdensome to the bacterial cell (San Millan and MacLean, 2017; Baltrus, 2013). That is, the plasmid costs may exceed the advantageous attributes harboured by the plasmid (Hall *et al*., 2015). In theory, several studies have postulated the instability of plasmids evolutionarily, regardless of whether the plasmid costs supercede its usefulness or vice versa (Macken *et al*., 2017; Lili *et al*., 2007; Bergstrom *et al*., 2000; Stewart and Levin, 1977). When the costs of the presence of the plasmid exceeds its benefits to the bacterial host, within a short period, negative selection often results. This is due to reduction in the number of the plasmids. Such a situation is difficult to prevent if there is low rates of conjugal gene transfers (Bergstrom *et al*., 2000; Simonsen, 1991). Hence, when there is negative selection, the persistence of the plasmid would depend on its capability to survive within the bacterial population. Plasmids can be maintained at high frequency by positive selection when the advantage of auxiliary genes exceeds the costs of harbouring the plasmid (Harrison *et al*. 2015; San Millan *et al*. 2014). It is however anticipated that their persistence would reduce over time because of the possibility that positive selection will enhance incorporation of useful auxiliary genes into the host chromosome (Bergstrom *et al*., 2000), a procedure promoted through auxiliary genes positioned on transposons or on any other integrative determinant (Osborn and Böltner, 2002).

Plasmids create multiple layered fitness cost to a bacterial cell (Baltrus, 2013). This is because of the negative effects produced from the burden placed by plasmid metabolism (Shachrai *et. al*., 2010; Bragg and Wagner, 2009), often originating from

the distinct regulatory associations among the bacterial cell and the plasmid (Baños *et. al.,* 2009; Doyle *et. al*., 2007), as well as the cytotoxic effects acquired from the plasmid presence in the host (Baltrus, 2013).

The use of bioinformatics tools has shown large genetic diversity within the three lineages of phylogroup B2 in terms of multi-resistance genes, virulence toxins, conjugative systems, incompatibility groups and integrons (chapter two). However, there is the need to understand if the plasmids observed among these lineages vary in their ability to transfer to a common host strain, if that variation depends on the ST lineage, the plasmid or both. There is also the need to understand how long these plasmids can stably persist within the lineages and the effects of their presence on the *E. coli* host. The aims of this project were thus, to determine the rate of plasmid transfers using plasmid transfer experiments, to examine the impacts of the plasmids on transfer rates using GWAS, to investigate the frequency in which plasmids are lost within the transconjugants strains of the phylogroup B2 lineages, since this will depend on how these plasmids are stably maintained within the bacterial population and finally, to examine the impacts of plasmid carriage on the host fitness.

3.3 Materials and Methods

3.3.1 Strain Collection

We assembled a collection of 180 *Escherichia coli* isolates of STs 73, 95 and 131, for which whole genome sequence data were available. The *E. coli* isolates where from blood, urine and faeces taken from people living in the Canberra region of Australia

(Gordon, *et al*., 2015; Blyton *et al*., 2014; Gordon, 2010; Gordon, *et al*., 2005; Gordon, *et al*., 2002). The choice of strains was based on strains with available MobSuite predictions, belonging to incompatibility group, IncF and resistant to either ampicillin or tetracycline. For the plasmid transfer experiments, strains to act as donors were resistant to ampicillin or \ and tetracycline while the recipient strain is a plasmid-free, nalidixic acid resistant, and lactose-negative laboratory strain of *E. coli* K-12 (CSH50) [A(lac-pro) ara str nal]. We were eventually left with a set of 93 strains that met these criteria. The strains were kept at -81^{0}C in 6.25% glycerol volume by volume (v/v) pure cultures. Plasmid stability tests were carried out by using sixty (60) transconjugants strains derived from the conjugal transfers while host fitness test was investigated by using same transconjugants strains and the plasmid-free ancestral *E. coli* K-12 recipient strain CSH50 from the conjugal transfers. The transconjugants were kept at -81^{0}C in 6.25% glycerol volume by volume (v/v) pure cultures.

3.3.2 Antimicrobial susceptibility testing of all strains

Antimicrobial profiles of the bacterial strains against the antibiotics ampicillin, tetracycline and nalidixic acid were confirmed by using Kirby-Bauer disc diffusion method on Muller-Hinton agar as described by the Clinical and Laboratory Standards Institute (CLSI, 2007). Freezer cultures were streaked onto MacConkey agar plates and incubated over the night at 37^{0}C. Single colonies from each of the plates were tooth picked into 5ml Lysogeny broth (LB) and incubated at 35^{0}C and 150 revolution per minute (rpm) in a shaker overnight. 100µl of the overnight culture was inoculated into 3ml soft agar (0.6% weight / volume in Miller's LB broth), vortexed and decanted onto the surface of Miller's LB agar plates. Filter discs that have been embedded with

antibiotics were placed on the surface of the plates. The plates were then incubated at 37^0C overnight. Zones of inhibition surrounding the antimicrobial discs denote the sensitivity of the strains to the selected antibiotics.

After the conjugal transfers, antimicrobial profiles of the donor conjugative *E. coli* cells and the transconjugants were determined using different antibiotics by the method of Kirby-Bauer disc-diffusion technique (Bauer *et al*., 1966). The following antibiotics representative of different classes were used: tetracyclines {tetracycline 30µg}, nitroimidazoles {metronidazole 80µg}, phenicol {chloramphenicol 30µg}, fluoroquinolones {ciprofloxacin 5µg}, trimethoprim {trimethoprim 5µg}, quinolones {nalidixic acid 30µg}, carbapenems {ertapenem 10µg}, aminoglycosides {gentamicin 10µg}, third-generation cephalosporins {cefotaxime 30µg}, penicillins {amoxicillin / clavulanic acid 20/10µg and ampicillin 10µg}, nitrofurans {nitrofurantoin 100µg} and sulphonamides {sulphafurazole 300µg}.

3.3.3 Antibiotics Preparation for conjugal transfers

Tetrazolium chloride lactose medium was prepared with appropriate antibiotics in the following concentrations: Ampicillin (Amp) in 200mg/ml of distilled water, tetracycline (Tet) 5mg/ml in ethanol, Nalidixic acid (Nal) 20mg/ml in 1N NaOH.

3.3.4 Bacterial conjugation

Conjugal transfers were carried out as detailed by Shohayeb *et al. (*1994): the donor and the recipient cells were inoculated separately in 10 ml non-selective Miller's LB broth (Bacto) and incubated overnight at 35^0C at 150rpm for 18 hr. The following day,

in an already labelled 10mls LB flasks, 0.1ml of each donor was added to 0.1ml of the recipient. This was done in replicates. The flasks were then incubated overnight in the shaker at 35^{o}C with 150 rpm. The next day, growth from the flasks were serially diluted in Eppendorf tubes containing 0.99ml saline solution. 0.1ml from the flask in the overnight growth culture, first growth diluent (10^{-2}), second dilution tube (10^{-4}) and the third dilution tube (10^{-6}) were plated separately and in replicates directly on prepared tetrazolium chloride / lactose plates containing ampicillin for donors, nalidixic acid for the recipient cell and combination of ampicillin and nalidixic acid for transconjugants (products of the conjugal transfers). All plates were incubated at 37^{0}C for 24hours.The culture growth from each flask, first and second dilution tubes were analysed for transconjugants while the culture growth from the third dilution tubes were analysed for the donor and the recipient cells. Resistance determinants were examined by evaluating all transconjugants against the antibiotics that the donor strains were resistant to. Single colony of each of the transconjugant was inoculated into 10mls non-selective Miller's LB (Bacto), and incubated overnight at 35^{0}C at 150 revolution per minute for 18 hr. The following day, 1ml of each transconjugant was stored at -81^{0}C in 6.25% glycerol v/v for stability test.

The rate of plasmid transfer was approximated using the following formula:

$$\propto \frac{\text{Transconjugant cell/ml}}{\text{Donor cell/ml x Recipient cell/ ml}}$$

3.3.5 Presence or absence of genes

The assembled genomes were annotated using Prokka v1.14.6 (Seemann, 2014). Pangenomes of the strains were analysed using Roary v3.13.0 (Page *et al*., 2015).

Roary generated clusters of similar gene groups and produced a matrix for the presence and absence of ortholog auxiliary genes.

3.3.5.1 Random Forest Method

Pangenome data were used to predict log average transconjugant density from the conjugal transfers using random forest method. The random forest is a machine learning ranking program, whose algorithm is composed of numerous decisions trees (Roguet *et al*., 2018; Breiman, 2001). The method uses ensemble learning, which is a technique that combines many classifiers to produce 'a forest of trees'. The 'forest' produced by random forest algorithm uses bagging or bootstrap aggregating when constructing each separate tree. The random forest algorithm uses the predictions of the decision trees to generate the required outcome (Roguet *et al*., 2018; Cutler *et al*., 2007; Breiman, 2001). It predicts the results by taking the mean of the output from various trees. Increasing the number of trees enhances the precision of the outcome.

3.3.6 Plasmid loss measurement

The plasmid populations were propagated in the absence of antibiotics for 7days. To test for plasmid stability, sixty (60) freezer cultures of *E. coli* transconjugant strains were streaked out on LB / ampicillin plates and grown at 37°C overnight. Single colonies from these plates were then subjected to the following assays for plasmid loss: a single colony from each plate was inoculated into 10ml LB without antibiotics and grown at 37°C with shaking at 150 rpm. These cultures were passaged in replicates every 24h by transferring 100µl each into 10ml of fresh LB without antibiotics. After seven passages, aliquot of 10µl from each culture was transferred

into saline tubes and were serially diluted in replicates. 100µl of the dilution was then plated on LB agar plates with and without antibiotic (ampicillin) in six replicates and incubated at 37°C overnight. The next day, 100 to 300 random colonies from each of these plates were counted to check for colonies that lost ampicillin resistance (indicating plasmid loss) and those that had maintained the plasmid. The rate in which plasmids were lost within the strains of the three lineages on the seventh (7th) day was noted. The stability frequency of the plasmids was calculated by using the following formula:

log10 (Ng) / log10 (Nw),

Where Ng represent densities of bacterial cells containing the plasmid and Nw represent all the bacterial cells in the medium.

3.3.7 Host fitness test

The competitive experiment was carried out by using transconjugant strains generated from conjugal plasmid transfers and the plasmid-free recipient K12 strain CSH50. For the competition experiments, relative fitness of plasmid carrying, and plasmid-free strain were estimated by removing aliquots of the transconjugants strains and the plasmid-free strain from the vials in which they were stored at −80^{0}C. They were streaked out on tetrazolium chloride lactose (TCL) plates and tetrazolium chloride/ampicillin plates for the K-12 strain and transconjugants strains respectively. The plates were incubated at 37°C overnight. Single colonies from these plates were then, inoculated into separate flasks containing lysogeny broth. The cultures were also

grown overnight in a shaker at 150 rpm and 35^0C till the stationary phase was reached. They were then mixed at a 1:1 ratio (100-µl each) in the competition environment. Initial densities at time point (t = 0), were estimated by taking an initial aliquot of 10µl from each of these cultures, transferred into saline tubes and serially diluted in replicates. Finally, 100-µl sample of each of this mixture was taken, spread onto prepared TCL plates and tetrazolium chloride/ampicillin plates and incubated at 35^0C for 24hours.

To obtain the cultures for the final growth, the competition mixture was incubated in the same conditions as the initial culture for 24 h. Also, 100-µL from the final sample, that has been serially diluted, was spread on TCL plates and tetrazolium chloride/ampicillin plates and incubated at 37^0C. The competitors on the plates were counted and their numbers were multiplied using the dilution factor to determine their initial and final population sizes. The strains were distinguished through plating of 100µl of the dilution on TCL plates (non –selective) and tetrazolium chloride/ampicillin (selective). Subtraction of colonies in non -selective plates from the selective plates gives the growth counts of the K-12 strains.

All cultures and platings were done in five replicates each. Colonies were calculated by counting viable cells at time points 0 and 24 h. From the initial and final population sizes of the strains, the relative fitness can be calculated using the following formula:

$$W = \frac{\log10 \{N_A(f) / N_A(i)\}}{\log10 \{N_B(f) / N_B(i)\}}$$

Where W is relative fitness,

$N_A(f)$ = cell density of the transconjugants at 24h.

$N_A(i)$ = cell density of the transconjugants at 0.

$N_B(f)$ = cell density of the recipients at 24h.

$N_B(i)$ = cell density of the recipients at 0.

3.3.8 Statistical Analysis.

All analyses were conducted in JMP 15 statistical package. Non-parametric tools such as Median test, Wilcoxon and van der Waerden Test were used to conduct one–way analysis with chi-square approximations for conjugal transfers. Analysis of data from the random forest method was done using bootstrap method that has been incorporated in JMP 15 software.

3.4 Results

3.4.1 Donor strains profile

All the strains that were resistant to either ampicillin or tetracycline and sensitive to nalidixic acid were used as donors. The plasmids encoding this resistance were found to be associated with specific F-plasmid sequence type (pMLST). The ST lineages, F-plasmid sequence type, Inc group and resistance patterns of each donor strain is shown (table 3.1).

Bacterial strain (donor)	Lineages	F-plasmid sequence type	Inc group	Donor resistance pattern
H249	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
BS076	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
BS058	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
BS150	ST95	F18:A-:B1	IncFII, IncFIB	Tetracycline

Bacterial strain (donor)	Lineages	F-plasmid sequence type	Inc group	Donor resistance pattern
BS157	ST95	F24:A-:B1	IncFII, IncFIB	Ampicillin
5-2-R8	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
H648	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
2H-327-14	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
H087	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
BS117	ST95	F18:A-:B1	IncFII, IncFIB	Tetracycline
H008	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
H219	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
M675220	ST95	F29:A-:B10	IncFII, IncFIB	Ampicillin
H144	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
M668546	ST95	F35:A-:B1	IncFII, IncFIB	Ampicillin
BS142	ST95	F2:A-:B1	IncFII, IncFIB	Ampicillin
BS020	ST95	F18:A-:B1	IncFII, IncFIB	Tetracycline
H552	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
H355	ST95	F51:A-:B10	IncFII, IncFIB	Ampicillin
45-1-Ti1	ST95	F29:A-:B10	IncFII, IncFIB	Ampicillin
40-1-R9	ST95	F29:A-:B10	IncFII, IncFIB	Ampicillin
39-1-AC10	ST95	F29:A-:B10	IncFII, IncFIB	Ampicillin
70-4-DC20	ST95	F29:A-:B10	IncFII, IncFIB	Ampicillin
2H-265-13	ST95	F18:A-:B1	IncFII, IncFIB	Ampicillin
H140	ST95	F18:A-:B1	IncFII, IncFIB	Ampicillin
H151	ST95	F24:A-:B1	IncFII, IncFIB	Ampicillin
H244	ST95	F107:A-:B1	IncFII, IncFIB	Ampicillin
H154	ST95	F24:A-:B1	IncFII, IncFIB	Ampicillin
BS045	ST95	F24:A-:B1	IncFII, IncFIB	Ampicillin
BS014	ST95	F2:A-:B1	IncFII, IncFIB	Ampicillin
BS011	ST95	F2:A-:B1	IncFII, IncFIB	Ampicillin
M638488	ST95	F24:A-:B1	IncFII, IncFIB	Ampicillin
M660195	ST73	F-:A-:B-	IncFIB	Ampicillin
M660331	ST73	F-:A-:B-	IncFIB	Tetracycline
M670800	ST73	F4:A-:B-	IncFII	Ampicillin
20-5-R7	ST73	F29:A-:B10	IncFII, IncFIB	Ampicillin
H003	ST73	F2:A-:B1	IncFII, IncFIB	Ampicillin
H118	ST73	F2:A-:B1	IncFII	Ampicillin
H048	ST73	F2:A-:B1	IncFII	Ampicillin
H274	ST73	F1:A1:B1	IncFIA, IncFII, IncFIB	Ampicillin
H078	ST73	F2:A-:B1	IncFII	Ampicillin
BS373	ST73	F2:A-:B1	IncFII	Ampicillin
BS486	ST73	F29:A-:B10	IncFII, IncFIB	Ampicillin
BS457	ST73	F29:A-:B10	IncFII, IncFIB	Ampicillin
H309	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
BS471	ST73	F22:A1:B20	IncFIA, IncFII, IncFIB	Ampicillin

Bacterial strain (donor)	Lineages	F-plasmid sequence type	Inc group	Donor resistance pattern
BS404	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
18-3_Ti5	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H378	ST73	F1:A1:B20	IncFIA, IncFII, IncFIB	Ampicillin
BS475	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H050	ST73	F51:A-:B10	IncFII, IncFIB	Tetracycline
H429	ST73	F2:A-:B1	IncFII, IncFIB	Ampicillin
50-M694910	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
BS478	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
M672775	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
M587611	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
M587973	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H608	ST73	F2:A-:B1	IncFII	Ampicillin
H554	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H187	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H169	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H361	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
H201	ST73	F1:A1:B1	IncFIA, IncFII, IncFIB	Ampicillin
2-1-P20	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
M619205	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
M630782	ST73	F51:A-:B10	IncFII, IncFIB	Ampicillin
M652957	ST131	F18:A-:B1	IncFII, IncFIB	Ampicillin
BS418	ST131	F4:A-:B52	IncFIA, IncFII, IncFIB	Ampicillin
BS470	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
BS003	ST131	F18:A-:B1	IncFII, IncFIB	Ampicillin
BS488	ST131	F1:A1:B1	IncFIA, IncFII, IncFIB	Ampicillin
M683442	ST131	F-:A-:B-	IncFIB	Tetracycline
M652483	ST131	F1:A1:B1	IncFIA, IncFII, IncFIB	Ampicillin
M653835	ST131	F29:A-:B10	IncFII, IncFIB	Tetracycline
M608681	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
M698766	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
M694204	ST131	F-:A-:B-	IncFIB	Tetracycline
M709320	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
M685551	ST131	F29:A-:B10	IncFII, IncFIB	Tetracycline
H090	ST131	F4:A-:B10	IncFII, IncFIB	Ampicillin
H407	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
H346	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
H606	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
H392	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
H183	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
M704989	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
M710059	ST131	F2:A-:B1	IncFII, IncFIB	Ampicillin
M698766	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin

Bacterial strain (donor)	Lineages	F-plasmid sequence type	Inc group	Donor resistance pattern
BS078	ST131	F1:A2:B20	IncFII, IncFIB, IncFIA	Ampicillin
BS377	ST131	F2:A-:B1	IncFII, IncFIB, IncFIA	Ampicillin
BS086	ST131	F2:A-:B1	IncFIA	Ampicillin
32-2-R12	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin
33-1-TC19	ST131	F29:A-:B10	IncFII, IncFIB	Ampicillin

Table 3-1 The profile for each of the donor strains among the ST lineages

3.4.2 Variation in the rate of plasmid transfers

This study shows that plasmid was transferred by all 93 donor strains of the lineages analysed, but the rate of transfer did not depend on the genomic background (ST) of the donor strains (fig. 3.1), (Median Test: X^2 =0.19, P = 0.90, P > X^2). There were some fairly small differences in the rate of plasmid transfer with the F- plasmid sequence types of the plasmid being transferred (fig. 3.2) (van der Waerden Test: X^2 = 5.1, P < X^2, P= 0.16), n= 79 for common F-plasmid sequence types F18:A-:B1, F2:A-:B1, F29:A-:B10 and F51:A-:B10 (fig. 3.2). The analysis was restricted to those F-plasmid sequence types with sample sizes greater than 6.

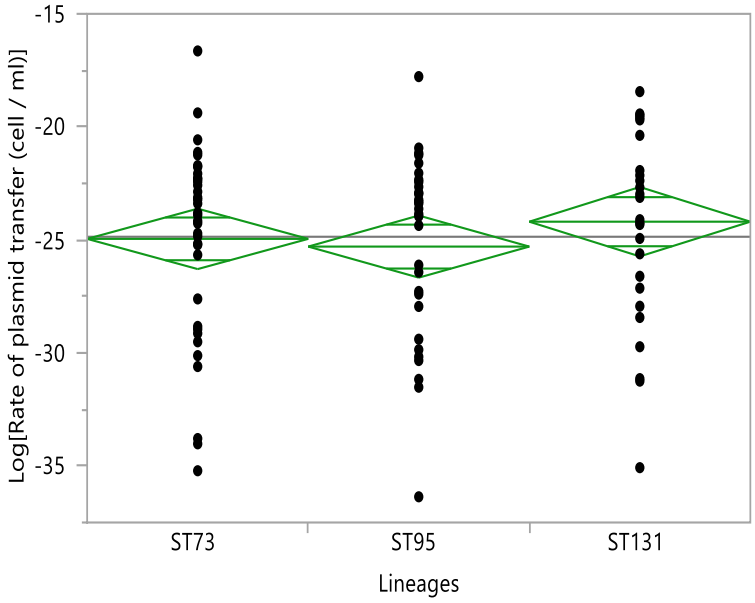

Figure 3.1 Rates of plasmid transfer by ST lineages

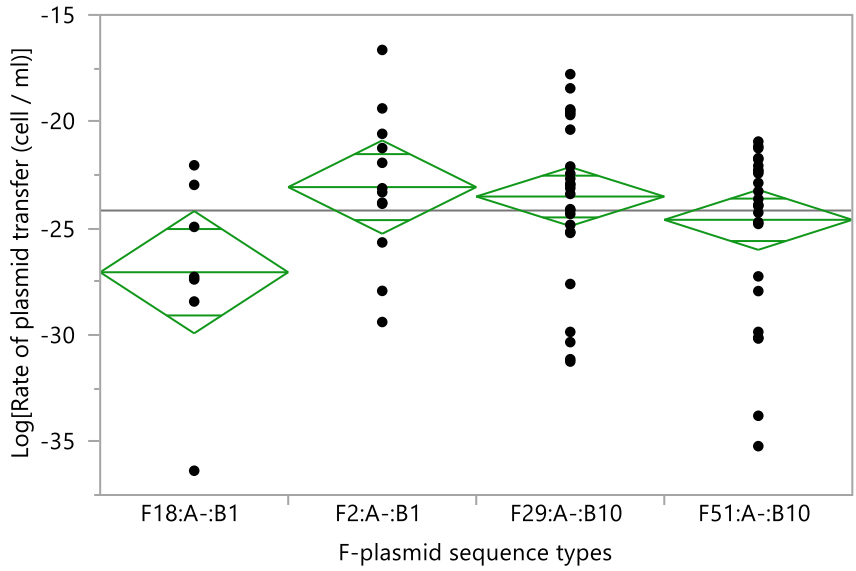

Figure 3.2 Rates of plasmid transfer by F-plasmid sequence types

3.4.3 Major auxiliary (*tra* genes) associated with plasmid transfer rate.

The random forest analysis was based on all strains analysed, but the results of the analysis were based on four major F-plasmid sequence types, F29:A-:B1, F51:A-:B10, F2:A-:B1 and F1:A1-:B1 shared among the strains of the three lineages. The other sequence types are either found only in one lineage or have a sample same less than two. Several genes were detected by the random forest method. Most of these genes are known as *tra* genes and are often found in the *tra* operon. The genes detected exhibited the greatest effect on the transfer rate (table 3.2).

Genes
Conjugal transfer protein - *traH*
Conjugal pilus assembly protein - *traF*
Conjugal transfer protein - *traR*
Type IV secretion system protein – *traC*

Tetracycline resistance - associated transcriptional repressor - *tetC*
Electron transport complex subunit - *rsxC*

Table 3-2 Some of the genes detected by random forest method

3.4.4 Frequency distribution of *tra* genes associated with key F-plasmid Sequence types

In this analysis, *tra* genes of interest are those encoded by F-plasmid sequence types F29:A-:B1, F51:A-:B10, F1:A1-:B1 and F2:A-:B1 within the strains of the three ST lineages. Specifically, the *tra* genes, *traF, traH, traR* and *traC* were numerous within the F-plasmid sequence types, F51:A-:B10 and F29:A-:B1. 100% of all *tra* genes detected were found in all ST73 and ST95 strains with F29:A-:B1. *TraH* and *traC* were also observed in all strains of ST131. F51:A-:B10 encoded 100% of most of the *tra* genes in ST95 strains and 98% for ST73. The *tra* genes were not prevalent in the other F-plasmid sequence types (fig. 3.3).

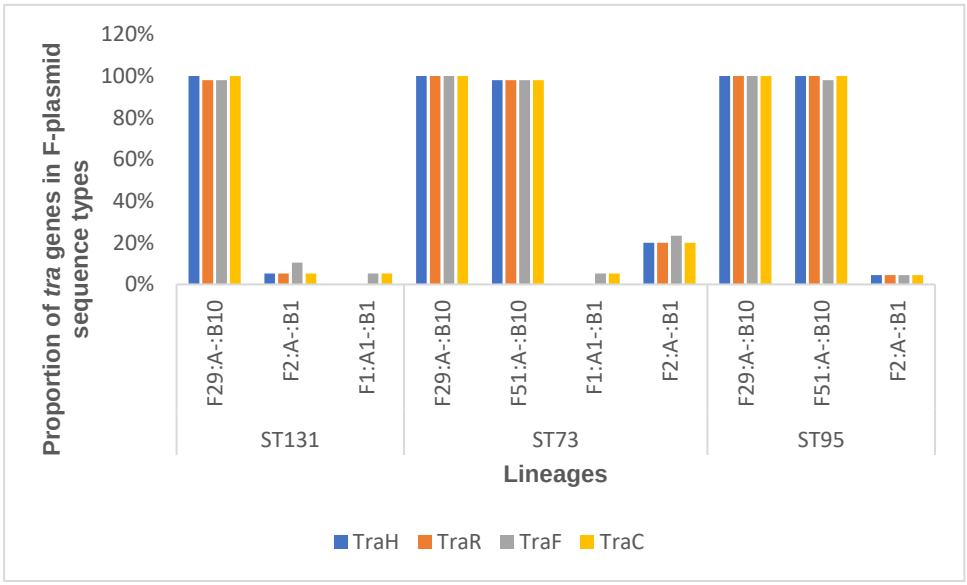

Figure 3.3 Distribution of each *tra* gene within the major F-plasmid sequence types.

3.4.5 Impacts of *tra* genes on plasmid transfers

tra genes found in the F-plasmid sequence types: F29:A-:B1 and F51:A-:B10 were observed to positively enhance the rates of plasmid transfers (figs. 3.4, 3.5, 3.6, and 3.7). In addition, significant transconjugant yields were found in strains with multiple *tra* genes *traH, traR* and *traF* (fig.3.8). Strains with two *tra* genes were not analysed due to their small number. In the figures displayed, level 1 represents genes that were transferred while level 0 represents the absence of *tra* genes. Log e scale was used for all the graphs.

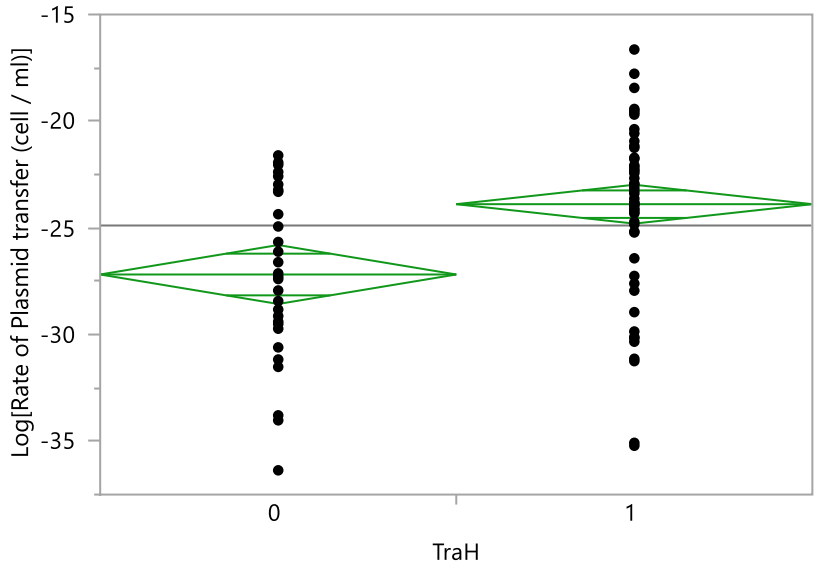

Figure 3.4 **Associated genes of *traH* gene on plasmid transfer (Wilcoxon test: $X^2 = 13.9$, $P < X^2$, $P = 0.0002$).**

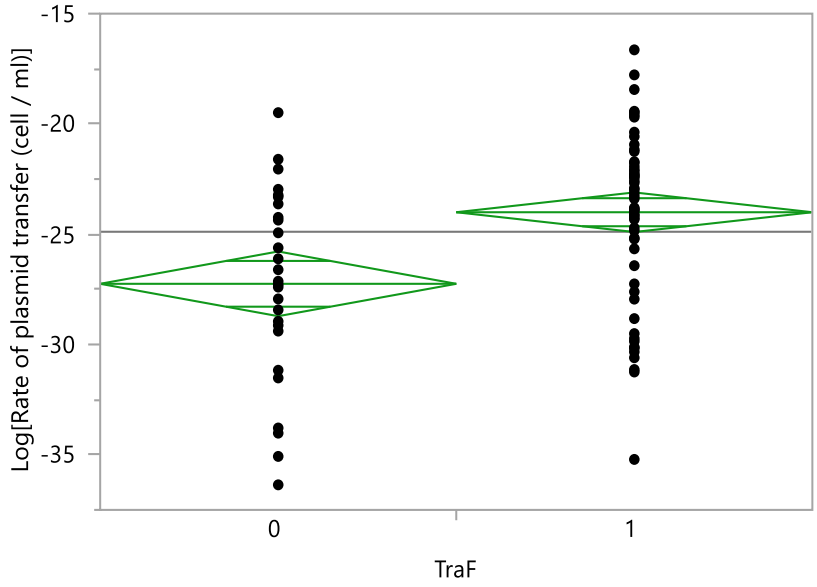

Figure 3.5 **Associated genes of *traF* gene on plasmid transfer (Wilcoxon test: X^2 = 11.9, P < X^2, P = 0.0005).**

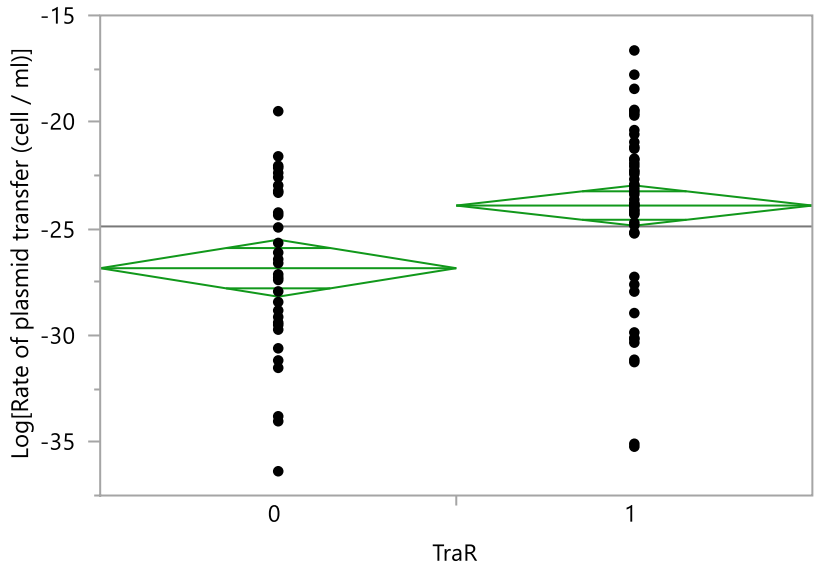

Figure 3.6 **Associated genes of *traR* gene on plasmid transfer (Wilcoxon test: X^2 = 12.4, P < X^2, P = 0.0004).**

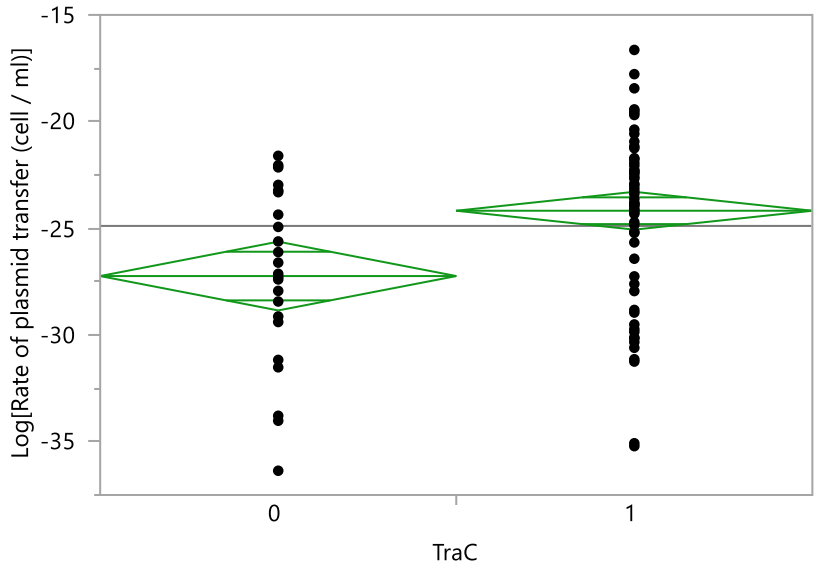

Figure 3.7 Associated genes of *traC* gene on plasmid transfer (Wilcoxon test: X^2 = 9.7, P < X^2, P = 0.002).

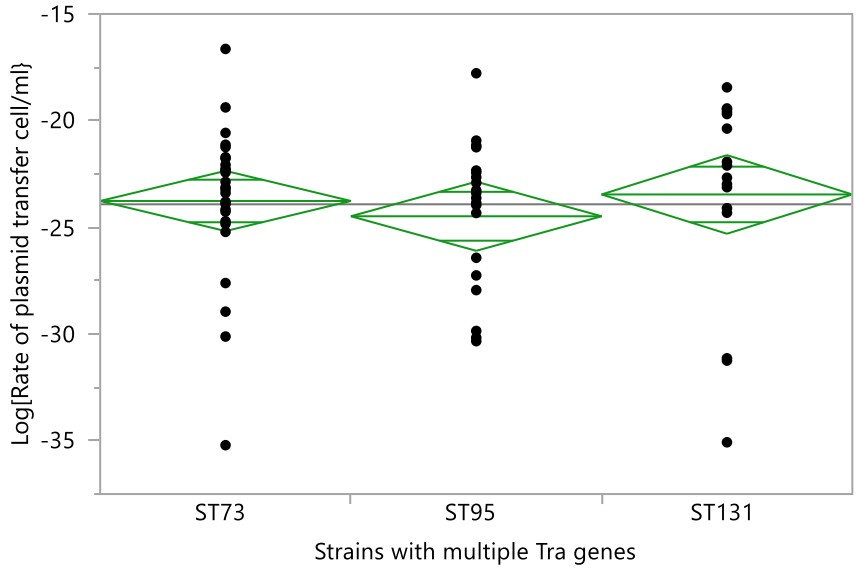

Figure 3.8 Associated genes of multiple *tra* genes on rate of plasmid transfer (Wilcoxon test: X^2 = 6.16, P < X^2, P = 0.05).

3.4.6 Antibiotic resistance genotypes of donor bacteria and transconjugants

The resistance gene phenotypes of transconjugants compared to their donor strains using Kirby-Bauer disc-diffusion technique are shown in table 3.3. The results show

that all the resistance determinants were successfully transferred during bacterial conjugation.

Bacterial strain (donor)	Lineages (STs)	Donor Resistance pattern	Selective media supplement		
			Donor	Recipient	Transconjugants resistance pattern
M587973	ST73	SF-AMP	AMP	Na	SF-AMP
M619205	ST73	SF-AMP	AMP	Na	SF-AMP
M630782	ST73	SF-AMP	AMP	Na	SF-AMP
M660195	ST73	SF-AMP	AMP	Na	SF-AMP
M660331	ST73	TE-SF	AMP	Na	TE-SF
M670800	ST73	SF-AMP	AMP	Na	SF-AMP
H169	ST73	SF-AMP	AMP	Na	SF-AMP
H003	ST73	SF-AMP	AMP	Na	SF-AMP
H078	ST73	SF-AMP	AMP	Na	SF-AMP
BS373	ST73	SF-AMP	AMP	Na	SF-AMP
BS471	ST73	SF-AMP	AMP	Na	SF-AMP
BS404	ST73	SF-AMP	AMP	Na	SF-AMP
18-3_Ti5	ST73	SF-AMP	AMP	Na	SF-AMP
H378	ST73	SF-AMP	AMP	Na	SF-AMP
BS475	ST73	SF-AMP	AMP	Na	SF-AMP
H050	ST73	TE-SF	AMP	Na	TE-SF
50-M694910	ST73	SF-AMP	AMP	Na	SF-AMP
BS478	ST73	SF-AMP	AMP	Na	SF-AMP
H608	ST73	SF-AMP	AMP	Na	SF-AMP
H309	ST73	SF-AMP	AMP	Na	SF-AMP
2-1-P20	ST73	SF-AMP	AMP	Na	SF-AMP
H048	ST73	AMP	AMP	Na	AMP
20-5-R7	ST73	AMP	AMP	Na	AMP
BS457	ST73	AMP	AMP	Na	AMP
H118	ST73	AMP	AMP	Na	AMP
H187	ST73	SF-AMP	AMP	Na	SF-AMP
H274	ST73	SF-AMP	AMP	Na	SF-AMP
H361	ST73	SF-AMP	AMP	Na	SF-AMP
H554	ST73	SF-AMP	AMP	Na	SF-AMP
H201	ST73	TE-SF-AMP-C	AMP	Na	TE-SF-AMP-C
M672775	ST73	SF-AMP	AMP	Na	SF-AMP
M587611	ST73	SF-AMP	AMP	Na	SF-AMP
M647320	ST73	AMP	AMP	Na	AMP
H429	ST73	AMP	AMP	Na	AMP
BS486	ST73	SF-AMP	AMP	Na	SF-AMP

Bacterial strain (donor)	Lineages (STs)	Donor Resistance pattern	Selective media supplement		
			Donor	Recipient	Transconjugants resistance pattern
H249	ST95	SF-AMP	AMP	Na	SF-AMP
BS076	ST95	SF-AMP	AMP	Na	SF-AMP
BS058	ST95	SF-AMP	AMP	Na	SF-AMP
BS150	ST95	TE-SF-AMP-C	AMP	Na	TE-SF-AMP-C
5-2-R8	ST95	AMP	AMP	Na	AMP
2H-265-13	ST95	AMP	AMP	Na	AMP
2H-327-14	ST95	AMP	AMP	Na	AMP
H087	ST95	AMP	AMP	Na	AMP
BS117	ST95	TE-SF-AMP-C	AMP	Na	TE-SF-AMP-C
BS142	ST95	AMP	AMP	Na	AMP
BS011	ST95	AMP	AMP	Na	AMP
BS014	ST95	AMP	AMP	Na	AMP
BS020	ST95	TE-SF	AMP	Na	TE-SF
BS045	ST95	AMP	AMP	Na	AMP
H552	ST95	AMP	AMP	Na	AMP
H355	ST95	AMP	AMP	Na	AMP
H151	ST95	AMP	AMP	Na	AMP
H154	ST95	TE	AMP	Na	TE
M668546	ST95	AMP	AMP	Na	AMP
H140	ST95	AMP	AMP	Na	AMP
H144	ST95	TE	AMP	Na	TE
H244	ST95	SF-AMP	AMP	Na	SF-AMP
H219	ST95	AMP	AMP	Na	AMP
70-4-DC10	ST95	AMP	AMP	Na	AMP
45-1-Ti1	ST95	AMP-SF	AMP	Na	AMP-SF
40-1-R9	ST95	AMP	AMP	Na	AMP
39-1-AC10	ST95	AMP	AMP	Na	AMP
H648	ST95	AMP	AMP	Na	AMP
H008	ST95	SF-AMP	AMP	Na	SF-AMP
BS157	ST95	AMP	AMP	Na	AMP
M638488	ST95	AMP	AMP	Na	AMP
M675220	ST95	AMP-TMP	AMP	Na	AMP-TMP
M685551	ST131	TE	Te	Na	TE
M653835	ST131	TE	AMP	Na	TE
BS048	ST131	TE	AMP	Na	TE
BS470	ST131	AMP	AMP	Na	AMP
M608681	ST131	AMP-C-TMP-SF	AMP	Na	AMP-SF-TMP-C
M709320	ST131	AMP	AMP	Na	AMP
M704989	ST131	AMP	AMP	Na	AMP
M698766	ST131	AMP-SF	AMP	Na	SF-AMP

Bacterial strain (donor)	Lineages (STs)	Donor Resistance pattern	Selective media supplement		
			Donor	Recipient	Transconjugants resistance pattern
H407	ST131	AMP	AMP	Na	AMP
H346	ST131	TE	AMP	Na	TE
H606	ST131	TE	AMP	Na	TE
H392	ST131	AMP	AMP	Na	AMP
H183	ST131	AMP	AMP	Na	AMP
32-2-R12	ST131	TE-SFM-AMP-TMP-SF	AMP	Na	TE-SF-AMP-TMP-SFM
33-1-TC19	ST131	TE-SFM-AMP-TMP-SF	AMP	Na	TE-SF-AMP-TMP-SFM
BS086	ST131	AMP-SF	AMP	Na	SF-AMP
M710059	ST131	AMP-SF	AMP	Na	SF-AMP
BS377	ST131	AMP-SF	AMP	Na	SF-AMP
M694204	ST131	TE	AMP	Na	TE
M652957	ST131	AMP-SF	AMP	Na	SF-AMP
BS418	ST131	AMP-SF	AMP	Na	SF-AMP
H090	ST131	AMP-SF	AMP	Na	SF-AMP
BS003	ST131	AMP-SF	AMP	Na	SF-AMP
BS488	ST131	AMP-SF	AMP	Na	SF-AMP
M683442	ST131	TE	AMP	Na	TE
M652483	ST131	AMP-SF	AMP	Na	SF-AMP

Table 3-3 Antimicrobial profiles of donor and transconjugant cells from conjugal transfers. AMP: Ampicillin, Chl: Chloramphenicol, Cip: Ciprofloxacin, Na: Nalidixic acid, Ca: Cefotaxime, GT, gentamicin; TE, tetracycline; AMC, amoxycillin clavulanic; TMP, trimethoprim; ETP, rtapenem; SF, sulphafurazole; CTX, cefotaxime; FN, nitrofurantoin.

3.4.7 Plasmid loss frequency in the lineages

The genetic background of the host cell maintaining the plasmids can influence plasmid evolution and maintenance This study shows that there was significant difference in the plasmid stabilities among the transconjugant strains of the lineages particularly between the lineages STs 73, n=24, 95, n=21 and their counterpart ST131, n=15 (Wilcoxon test: X^2= 25.06, P<0.0001) (fig. 3.9). Frequency of plasmid loss was the same for all K12 plasmids of ST73 and ST95. Analysis of plasmid stability in relation to F-plasmid sequence types reveals differences in the plasmid loss within the plasmid types F51:A-B10, n=34 and F29:A-:B10, n=26 (Wilcoxon test: X^2= 16.71,

P<0.0001) (fig. 3.10). It is pertinent to note that F51:A-B10 is found only in the strains of ST73 and ST95 lineages. Although, F29:A-:B10 is also found in the strains of these two lineages, previous analysis (chapter 2) showed that this F- plasmid sequence type is predominantly encountered in most strains of ST131.

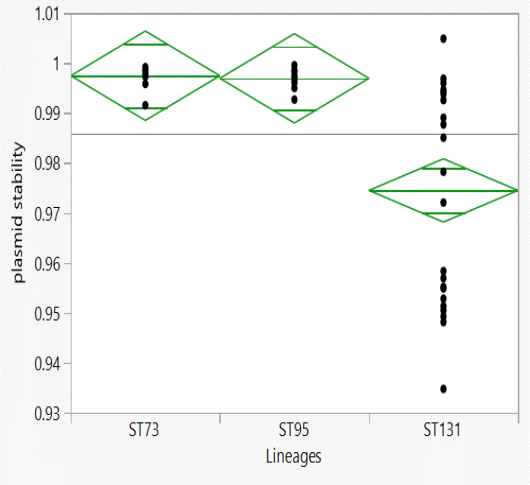

Figure 3.9 Plasmid stability rate by ST lineages

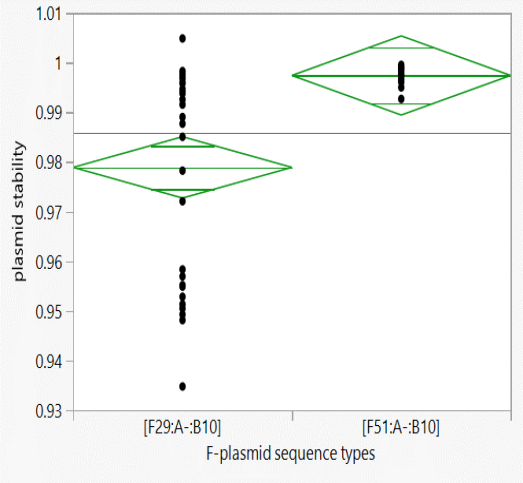

Figure 3.10 Plasmid stability rate by F-plasmid sequence types

3.4.8 Impacts of the plasmids on host fitness

There was a fairly small significant difference between the bacterial host fitness and the plasmids diversities among the transconjugant strains of the ST lineages ST73, n=24, ST95, n=21, ST131, n=15 (fig. 3.11) (Wilcoxon / Kruskal-Wallis Tests: $X^2 = 8.67$, P< 0.013).

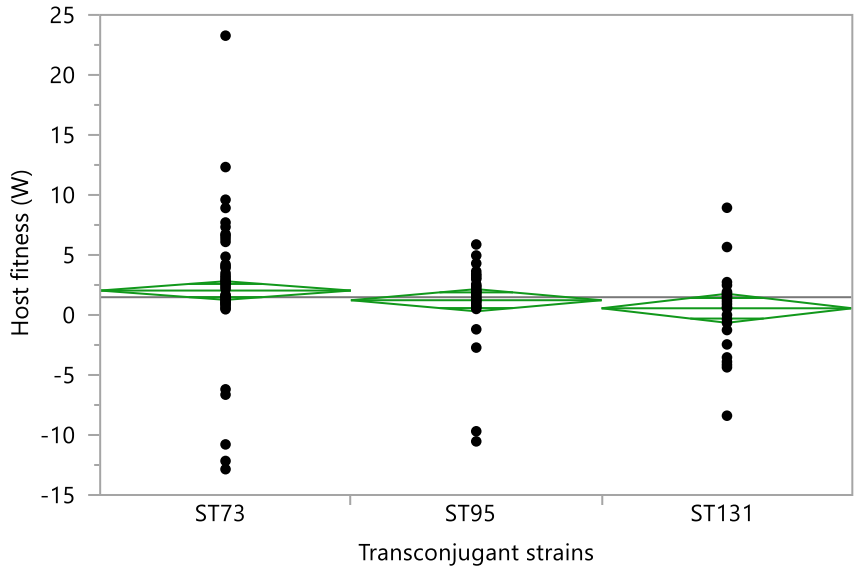

Figure 3.11 Cost effects of the plasmids on the host

3.4.9 Impacts of F-plasmid sequence types on host fitness

There was significant difference between F-plasmid sequence types diversities, F29:A-:B10, n=26, F51:A-:B10, n=34 among transconjugant strains of the three ST lineages and the host fitness (fig. 3.12) (Wilcoxon / Kruskal-Wallis Tests: $X^2 = 7.63$, $P < 0.005$).

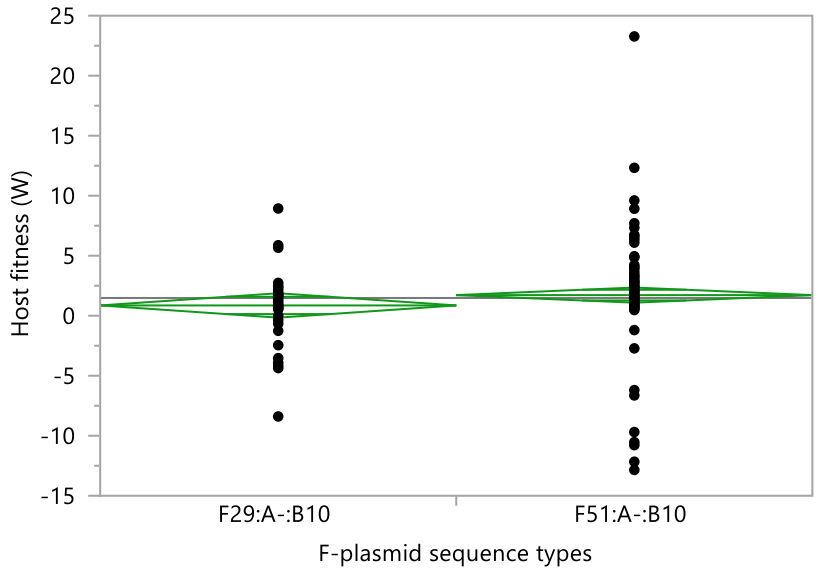

Figure 3.12 plasmid sequence types effects on host fitness

3.5 Discussion

The rate at which plasmids are transferred in lineages is one of the pivotal predictors of plasmid stability within the host (Stewart and Levin, 1977). Despite that all strains analysed through conjugal transfers were able to transfer their plasmids, the host genetic backgrounds have no effect on the rates of plasmid transfer. Although fairly small differences in plasmid transfer rates were observed within the F-plasmid sequence types, no significant difference was detected. However, the random forest analysis showed that the small differences observed in rates of plasmid transfer in relation to the F-plasmid sequence types could be explained by the presence / absence of certain genes mostly related to the *tra* operon. These genes such as conjugal transfer gene – *traH* and *traR*, conjugal pilus assembly gene – *traF,* type IV secretion system gene - *traC,* tetracycline resistance-associated transcriptional repressor -*tetC* and electron transport complex subunit –*rsxC, are* variable genes that play important roles in plasmid transfer. Most of these genes are *tra* genes found in the *tra* operon, where they play significant roles in bacterial conjugation (Lawley *et al*., 2003; Frost *et al*., 1994). The F-plasmid *tra* operon harbours many of the proteins that are needed for bacterial conjugal transfers (Kurenbach *et al.,* 2002). It is about ~30 kb and encodes approximately 40 genes. Each of these genes codes for a different protein product (Osbourn and Field, 2009; Lawley *et al*., 2003; Gubbins *et al*., 2002). The functions of these *tra* genes are diverse and interconnected to one another (Grohmann *et al.,* 2003). The *traC* is involved in DNA transfer. This gene often attach to the surface of conjugative plasmids and initiate conjugation. The genes*, traF, traH* and *traC* are responsible for pili assembly and production. These genes enable the formation and attachment of pilus to cell surfaces. They also function as periplasmic proteins and enhance the regulation of different regions of the *tra* operon itself. This latter function

ensures that organisms regulate the expression of a variety of genes depending on the conditions of the environment (Grohmann *et al.,* 2003*;* Clewell *et al*., 1993). This study further revealed that, even though some plasmids were still transferred in the absence of multiple *tra* genes, the presence of these *tra* genes in most of the plasmids, enhanced positively the rates of plasmid transfer.

The capacity of plasmids to undergo conjugal transfers is a key criterion in plasmid maintenance in bacterial populations. This is particularly important when they are able to transfer resistance determinants within the host communities. The knowledge of the factors that promote plasmid stability is very essential to studying bacterial evolution (Harrison and Brockhurst, 2012; Svara and Rankin, 2011; Lili *et al*., 2007; Bergstrom *et al*., 2000; Stewart and Levin, 1977). To study and compare the frequency of plasmid loss within the transconjugant strains of the phylogroup B2 lineages, we compared the evolution of the plasmid types found in the transconjugant strains of the three phylogroup B2 lineages by assessing the frequency of their plasmid stability using serial dilution transfer techniques. Frequency of plasmid loss was determined by investigating the plasmid persistence over a seven-day period. All plasmids in the transconjugant strains (which is now same as the K12 strain) evolved in the absence of antibiotics. Transconjugants populations from the ancestral strains of the lineages ST73 and ST95 that evolved were able to maintain a higher proportion of their plasmid when compared with some strains from ST131. This is because the frequency at which some plasmids evolved in the culture declined significantly after the 7^{th} day period in ST131 lineage. Our results reveal that the absence of antibiotics in the growth culture of the transconjugant strains of ST131 led to purifying selection against its' unstable plasmids resulting in decline over time. This situation did not take place with the

transconjugants strains of ST73 and ST95 and their stable plasmids were able to persist. But then, high frequency of plasmids are not needed to be present in order to have an effect on bacterial evolution especially in transient diverse environments, this is because even negligible plasmids can enhance the susceptibility of bacterial communities to fluctuating and unpredictable situations (Heuer and Smalla, 2012). Plasmid stability has been shown to confer plasmid persistence over an evolutionary timeline. But differences were also observed in plasmid stabilities in relation to F-plasmid sequence types of these transconjugants (Wilcoxon test: $X2= 16.7$, $P<0.0001$). This suggests that the variations in the plasmid stabilities in this study is not only influenced by the host genetic background but also on the type of F-plasmid sequence types in these plasmids.

The presence of plasmid within the host depend on the fitness costs of the plasmid to the host, and the benefits of the plasmids to the bacterial cell. The impacts of the plasmid carriage on the host is therefore essential because of its significance on plasmid adaptation and evolution. Fitness cost produced by the presence of plasmids often influence the host growth rate and reproductive mechanisms (Vogwill and MacLean, 2015) and the overall stability of the encoded plasmids. This fitness cost is a significant factor in the population biology of plasmids because the cost enhances selection against strains encoding plasmids. The significant differences observed in the fitness of the K12 host and the plasmids diversities among the transconjugant strains of the three ST lineages, as well as the F-plasmid sequence types diversities and the bacterial host fitness also reveal that host fitness cost can be impacted by both the host genetic background and plasmid encoded F-plasmid sequence type.

This study has shown that the evolution and dissemination of plasmids in clinical settings may be influenced by the genetic background of the host cell maintaining them and plasmid associated F-plasmid sequence types. The acquisition of resistance genes, virulence factors, integrons, bacteriocins, by the plasmids of these lineages (chapter 2), and numerous *tra* genes associated with the F-plasmid sequence types ensured large plasmid DNA contents, and consequently host determined plasmid-size limitations leading to the differences in plasmid stability and host fitness observed.

3.6 Conclusion

There is no variation in the rate of transfer of these plasmids with respect to the lineage of the bacterial host (STs). However, fairly small differences observed in rate of plasmid transfer in relation to F-plasmid sequence types reveals the presence / absence of genes in the *tra* operon. These *tra* genes were found mostly distributed within the two common F-plasmid sequence types F29:A-:B10 and F51:A-:B10. The unique functions of these *tra* genes are very essential in bacterial conjugation. Even though, the plasmid transfer rates is not dependent on the bacterial host, the host genetic background and the plasmid encoded F-plasmid sequence types may play important role in plasmid loss frequency and the host fitness .

3.7 References

Baltrus, D. A. (2013). Exploring the Costs of Horizontal Gene Transfer. *Trends Ecol Evol* **28**, 489–495.

Baños, R. C., Vivero, A., Aznar, S., García, J., Pons, M., Madrid, C. and Juárez, A. (2009). Differential Regulation of Horizontally Acquired and Core Genome Genes by the Bacterial Modulator H-NS. *PLoS genetics* **5**, e1000513.

Bauer, A. W., Kirby, W. M., Sherris, J. C. and Turck, M. (1966). Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* **45**, 493-6.

Bergstrom, C. T., Lipsitch, M. and Levin, B. R. (2000). Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155,** 1505-1519.

Blyton, M. D., Banks, S. C., Peakall, R., Lindenmayer, D. B. and Gordon, D. M. (2014). Not All Types of host contacts are equal when it comes to *E. coli* transmission. *Ecol Lett* **17**, 970–978.

Bragg, J. G. and Wagner, A. (2009) Protein Material Costs: Single Atoms can make an Evolutionary Difference. *Trends Genet* **25**, 5-8.

Breiman, L. (2001). Random forests. *Mach Learn* **45**, 5-32.

Clewell. D. B. (1993). In Bacterial Conjugation. Plenum Press, New York. *ISBN978-0-306-44376-3.*

Committee for Clinical laboratory Standards (2007). Performance standards for Antimicrobial Suspectibility testing: 17th International Supplement M100-S17. Wayne, PA: CLSI.

Cutler, D. R., Edwards, T. C. J., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology* **88,** 2783–92.

De Gelder, L., Ponciano, J. M., Joyce, P. and Top, E. M. (2007). Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology* **153**, 452–463.

Doyle, M., Fookes, M., Ivens, A., Mangan, M. W., Wain, J. and Dorman, C. J. (2007). An H-NS-like stealth protein aids horizontal DNA transmission in Bacteria. *Science* **315**, 251–252.

Frost, L. S., Ippen-Ihler, K. and Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**, 162–210.

Frost, L., Leplae, R., Summers, A. and Toussaint, A. (2005). Mobile Genetic Elements: The Agents of Open-Source Evolution. *Nat Rev Microbiol* **3**, 722–732.

Gordon, D. M. (2010). Strain Typing and the Ecological Structure of *Escherichia coli*. *J AOAC International* **93**, 974–984.

Gordon, D. M., Bauer, S. and James R. Johnson, J. R. (2002). The genetic structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiology*. **148**, 1513–1522.

Gordon, D. M., O'Brien, C. L. and Pavli, P. (2015). *Escherichia coli* Diversity in the Lower Intestinal Tract of Humans. *Environ Microbio Reports* **7**, 642–648.

Gordon, D. M., Stern, S. E. and Colington, P. J. (2005). Influence of the Age and Sex of Human Hosts on the Distribution of *Escherichia coli* ECOR Groups and Virulence Traits. *Microbiology* **151**, 15–23.

Grohmann, E., Muth, G. and Espinosa, M. (2003). " Conjugative Plasmid Transfer in Gram-Positive Bacteria" *Microbiology and Molecular Biology Reviews* **67,** 277–301.

Gubbins, M. J., Lau, I., Will, W. R., Manchak, J. M., Raivio, T. L. and Frost, L. S. (2002). The positive regulator, TraJ, of the *Escherichia coli* F plasmid is unstable in a cpxA* background. *J Bacteriol*. **184**, 5781-5788.

Hall, J. P. J, Harrison, E., Lilley, A. K., Paterson, S., Spiers, A. J. and Brockhurst, M. A. (2015). Environmentally co-occurring mercury resistance plasmids are genetically and phenotypically diverse and confer variable context-dependent fitness effects. *Environ Microbiol* **17**, 5008–5022.

Hall, J. P. J, Wood, A. J., Harrison, E. and Brockhurst, M. A. (2016). Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proc Natl Acad Sci USA* **113**, 8260-8265.

Harrison, E. and Brockhurst, M. A. (2012). Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol* **20**, 262–267.

Harrison, E., Guymer, D., Spiers, A.J., Paterson, S., and Brockhurst, M.A. (2015). Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr Biol* **25**, 2034–2039.

Heuer, H. and Smalla, K. (2012). Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev* **36**, 1083–1104.

Jain, R., Rivera, M. C., Moore, J. E. and Lake, J. A. (2003). Horizontal gene transfer Accelerates genome innovation and evolution. *Mol Biol Evol* **20**, 1598–1602.

Kurenbach, B., Grothe, D., Farías, M. E., Szewzyk, U. and Grohmann, E. (2002). The tra region of the conjugative plasmid pIP501 is organized in an operon with the first gene encoding the relaxase. *Journal of bacteriology* **184**, 1801–1805.

Lawley, T. D., Klimke, W. A., Gubbins, M. J. and Frost, L. S. (2003). F-factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**, 1-15.

Lili, L. N., Britton, N. F. and Feil, E. J. (2007). The persistence of parasitic plasmids. *Genetics* **177**, 309–405.

Macken, C. A., Levin, S. A. and Waldstatter, R. (2017). The dynamics of bacteria-plasmid systems. *J Math Biol* **32**, 123–145.

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000). Lateral gene transfer and the Nature of bacterial innovation. *Nature* **405**, 299–304.

Osborn, A. M. and Böltner, D. (2002). When Phage, Plasmids and Transposons Collide: Genomic Islands and Conjugative-and Mobilizable-Transposons as a Mosaic Continuum. *Plasmid* **48**, 202–12.

Osbourn, A. E. and Field, B. (2009). Operons. *Cellular and molecular life sciences* **66**, 3755- 3775.

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**, 3691–3.

Roguet, A., Eren, A. M., Newton, R. J. and Mclellan, S. L. (2018). Faecal source Identification using random forest. *Microbiome* **6,** 185.

San Millan, A., Heilbron, K., and MacLean, R.C. (2014). Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *ISME J* **8**, 601–612.

San Millan, A. and MacLean, R. C. (2017). Fitness costs of plasmids: a limit to plasmid transmission. *Microbiol Spectr* **5**, 1–12.

San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z. V., McLean, A. R., Cooper, B. S. and MacLean, R. C. (2014). Positive Selection and Compensatory Adaptation Interact to Stabilize Non-Transmissible Plasmids. *Nat Commun* **5**, 5208.

Seemann T. (2014). Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **30**, 2068–9.

Shachrai, I., Zaslaver, A., Alon, U. and Dekel, E. (2010). Cost of Unneeded Proteins in *E. coli* is Reduced after Several Generations in Exponential Growth. *Mol Cell* **38**, 758–767.

Shohayeb, M. and Sonbol, F. (1994). Conjugal transfer of antibiotic resistance of *Shigella* isolated in Tanta. *Alex J Pharm Sci* **8**, 177-80.

Simonsen, L. (1991). The Existence Conditions for Bacterial Plasmids: Theory and Reality. *Microb Ecol* **22**, 187–205.

Slater, F. R., Bailey, M. J., Tett, A. J. and Turner, S. L. (2008). Progress Towards Understanding the Fate of Plasmids in Bacterial Communities. *FEMS Microbiol Ecol* **66**, 3–13.

Stewart, F. M. and Levin, B. R. (1977). The population biology of bacterial plasmids: a priori conditions for the existence of conjugationally transmitted factors. *Genetics* **87**, 209–228.

Svara, F. and Rankin, D. J. (2011). The evolution of plasmid-carried antibiotic resistance. *BMC Evol Biol* **11**, 130.

Thomas, C. M. and Nielsen, K. M. *(*2005*).* Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria*. Nature Reviews Microbiology* **3***,* 711–721*.*

Uma, B., Prabhakar, K., Rajendran, S., Kavitha, K. and Sarayu, Y. L. (2009). Antibiotic Sensitivity and Plasmid Profiles of *Escherichia coli* Isolated from Pediatric Diarrhea. *Journal of global infectious diseases* **1**, 107–110.

Vogwill, T. and MacLean, R. C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol Appl* **8,** 284–295.

Chapter 4 Genomic structure comparison and phylogenetic analysis of the two common plasmid types using long-read data

4.1 Abstract

Compared to short-read sequencing, long-read sequencing enables effective and efficient investigation of the evolutionary history existing among different lineages. This is relevant because of the many biological changes that genomes often undergo during evolution, which often result in difficulty in characterising and comparing newly acquired sequences. Long-read sequence data of plasmids of the three phylogroup B2 lineages: ST131, ST73 and ST95 were analysed using various sequence comparison methods. Phylogenetic relationships existing among these plasmids were also investigated using various bioinformatics tools. This study shows that the sequencing method used played a central role in the number of plasmids, and plasmid DNA content observed in the genomes. Long-read results reveal that about 33% of all the assemblies harbour a single conjugative plasmid. Although, MobSuite underestimated the plasmid DNA in STs 131 and 95, it overestimated the number of plasmids in the three lineages compared to long-read. In addition, the study shows an ST73 effect on MobSuite predictions for plasmid number. Bandage analysis shows that about 56% of genomes from all three lineages analysed encode the same plasmid. Genomic comparisons performed across all genomes revealed that IncFII and IncFIB replicons were common among all genomes. These replicons were found to be associated with resistance genes, β-lactamase, *blaTEM-1B,* sulphonamide*, sul1* and aminoglycosides gene*, aadA1,* and is common in most plasmids of ST73 and ST95 lineages. In the plasmids of ST131, these same replicons were observed with plasmid

encoded virulence genes, enterobacterial complement resistance gene *(traT*) and secreted enterotoxin TieB, *senB*. *SenB* and *traT* were common across all genomes. These genes are very important in bacterial conjugal transfers. Genomic structure comparison performed using mauve software, and the phylogenetic tree analysis also revealed highly homologous structures, and similar evolutionary history among the two common plasmid types. It is however interesting to note that where similar plasmids were detected, they were found associated with a particular F-plasmid sequence type. F29:A-:B10 was common in most plasmids of ST131, although this F-plasmid sequence type was also detected in some plasmids of lineages ST73 and ST95, F51:A-:B10 is encoded in most of their genomes. Analysis of the gene content of the two common F-plasmid sequence types revealed that all the plasmids harboured about 74% of most of the *tra* genes that are involved in IncF plasmid conjugative transfer. and genes required for plasmid segregation and regulation, *parM* and *parA* were found in F29:A-:B10 and F51:A-:B10 plasmids respectively. Toxin-antitoxin genes such as *hokB*, *ccdA* and *ccdB were* over-represented in F29:A-:B10 plasmids while *peml* and *chpB* were common in F51:A-:B10 plasmids. These genes are necessary for the plasmid propagation and maintenance. Genes that regulate plasmid encoded virulence genes, *virB* and *virF* were detected only in F51:A-:B10 plasmids. Overall, this analysis reveals that plasmids of the two F-plasmid sequence types harbour diversity of different gene contents, whose functions may be inter-related and play multiple roles in the adaptation and evolution of the three *E. coli* lineages.

4.2 Introduction

During evolution, genomes are subjected to numerous small but extensive changes. These changes are often wide-ranging evolutionary processes that pose a threat to sequence comparison (Darling *et al*., 2004). Accurate understanding of many of these biological activities is more appreciated by investigating the frequency, trends, and patterns of each change type (Darling *et al*., 2004). Description of the frequency, patterns and trends of genome evolution was made possible by the emergence of data generated from genome sequencing (Darling *et al*., 2010; Darling *et al*., 2004).

Long read sequence data can be used to examine the evolutionary history existing among different plasmids through processes of genome comparisons, alignments, and phylogenetic tree building. Genome comparisons detect the presence of the same plasmids within different genomes. Genome alignments is used as a means of comparison by recognising evolutionary transitions within the DNA and aligning corresponding regions of sequence (Morrison, 2006). This process ensures the establishment of likely characteristics that are shared among progenies, thereby increasing our understanding of the sequences that have remained protected from those that have changed. Phylogenetic analysis is based on the use of detectable features of current plasmids to try to infer the chronology of occurrences that took place during the lineage of these plasmids (Morrison, 2011). Phylogenetic trees enhance our understanding of biodiversity and describe how closely related these plasmids are to one another. Similarities and differences in the plasmids can be used to deduce how closely related the lineages are, based on whether regions have changed, as well as how likely specific regions are to change. Ancestors can be deduced because their progenies have common distinct features.

Since the two common F-plasmid sequence types of the three phylogroup B lineages, play key roles in the bacterial conjugation, plasmid stability, host fitness, and overall, the adaptation and evolution of these lineages, in this analysis, the relationship existing in the two common F-plasmid sequence types was investigated using sequence comparisons methods and phylogenetic analysis for molecular sequence alignment and tree construction.

4.3 Materials and Methods

4.3.1 Source of bacterial strains and plasmids

Twenty-seven (27) long-read nanopore spade assemblies, derived from the analysis of the genomes of the lineages: STs 131, 73 and 95 were used for this study. The assemblies were from plasmids derived from *E. coli* strains bearing plasmids obtained from previously isolated clinical strains from Canberra hospital, Australia. The strains were selected based on preliminary data which suggested that the two pMLST types F29:A-:B10 and F51:A-B10 were shared among lineages. These strains were sequenced using minion system. Assembler Unicycler was used for assembly based on both Minion long read and illumina short read data.

4.3.2 Bandage analysis

Bandage is a bioinformatics application tool for navigating *De novo* assembly graphs. It is used to visualise links between assembled contigs (Wick *et al*., 2015). Each assembled genome was uploaded in graph fragment assembly (GFA) format within the bandage tool. The bandage tool was used to determine the number of plasmids in

a strain, their size, and total amount of plasmid DNA present. Fasta files of the plasmids selected were uploaded into basic local alignment search tool (nucleotide BLAST) (Altschul *et al*., 1990), to check the contigs derived plasmid. The blast hit was noted. Plasmids of interest were determined based on the presence of F-plasmid sequence types, F29:A-:B10 and F51:A-B10, presence of *tra* genes (protein BLAST), and incompatibility groups IncF.

4.3.3 Comparison of Long read results and MobSuite predictions

The long-read data were obtained through long read sequencing using nanopore assemblies while MobSuite predictions were from short read sequencing using SPAdes assemblies. Results obtained from the analysis of the long-read data was compared with the MobSuite predictions.

4.3.4 Characteristics of selected plasmids using genomic comparisons

4.3.4.1 Genomic features

Genomic features associated with plasmids of the phylogroup B2 lineages were compared using PlasmidFinder (Carattoli *et al*., 2014), MobileElementFinder (Johansson *et al*., 2020), antibiotic resistance (AR) genes, ResFinder 3.1 (Zankari *et al*., 2012), VirulenceFinder 2.0 (Joensen *et al*., 2014), The tools are found at the Centre for Genomic Epidemiology (CGE), Genomicepidemiology.org.

4.3.4.2 Rapid annotations subsystems technology (RAST) analysis

RAST is used to annotate or re-annotate the genomes of bacteria or archaea. It provides an environment for browsing annotated genomes, and then compare with hundreds of genomes maintained in the SEED integration (Overbeek *et al*., 2014). To analyse and compare plasmid content of the strains of the three lineages, the assembled genomes were uploaded to RAST server (Aziz *et al*., 2008). In this study, EC958 was used as the reference plasmid to compare the annotated genomes of all the plasmids, using protein BLAST with a cut off of 90% amino acid similarity across 90% of the predicted protein.

4.3.4.3 Mauve system

The Mauve system is used to construct multiple genome alignments of sequenced genomes. Here, two or more genomes that have undergone rearrangements due to recombination can be align (Darling *et al*., 2004). Multiple genome alignments provide a basis for research into comparative genomics and the study of genome-wide evolutionary dynamics. In this analysis, genomic structures of sequenced assemblies were explored and compared using progressive alignment algorithm in Mauve (version 2.3.1).

4.3.4.4 Phylogenetic analysis of the Sequence assemblies

Phylogenetic analysis involves three basic steps: 1) sequence alignment, this is used to find variable regions within the genomes, 2) character coding, necessary where there are gaps during sequence alignment, or the alignment were ambiguously aligned across the dataset (Morrison, 2011) while 3) tree building is used to unveil information

obtained from sequence alignment and the coding steps (Morrison, 2011). To determine the genetic relatedness of plasmids within strains of the three phylogroup B2 lineages, fasta files from long-read sequencing data of conjugative plasmids from the three lineages were uploaded to PhyML software (Guindon and Gascuel, 2003), to construct the phylogenetic tree. Bootstrap confidence values for each node of the trees were calculated over 100 replicate trees. Blast was used to confirm that the sequences are homologues, that is, they are a group of similar DNA sequences that share a common ancestry.

4.3.4.5 Analysis of the gene content of key F-plasmid sequence type

Long-read plasmid data were further analysed using Roary application tool (Page *et al*., 2015). Roary is used to produce clusters of similar gene groups.

4.3.5 Statistical analysis

Analyses were conducted in JMP 15 statistical package. Graphical representation was done in Excel package and Figtree application tool.

4.4 RESULTS

Long read sequencing was performed on 27 strains based on their genetic similarities: they all possess conjugative plasmids, have either F29:A-:B10 or F51:A-B10 sequence types, and have at least one integron each with specific resistance and virulence gene.

4.4.1 Long read results versus MobSuite predictions

The results from the Unicycler assemblies show complete sequence of the plasmid, but the results from MobSuite revealed short plasmid lengths because the plasmids did not assemble fully. There are differences in the number of plasmid, and plasmid DNA content in each assembly for the long-read results compared to MobSuite predictions (table 4.1). However, MobSuite results can predict the conjugation, mobilisable and non-mobilisable potential of each assembly. Although MobSuite underestimated the plasmid DNAs in STs 131 and 95 (fig. 4.1), it overestimated the number of plasmids in the three lineages. In addition, the results suggest an ST73 effect on MobSuite predictions for estimated number of plasmid (fig. 4.2).

Strain	ST Lineages	Long read results		MobSuite predictions	
		Total number of plasmids	Total plasmid DNA content	Total number of plasmids	Total plasmid DNA content
M173503	ST131	1	114kb	4	109kb
H471	ST131	1	114kb	3	111kb
BS411	ST131	2	188kb	2	128kb
55-1-AU4	ST131	4	228kb	4	114kb
M685551	ST131	2	125kb	2	111kb
M653842	ST131	5	138kb	8	157kb
M653835	ST131	1	114kb	3	113kb
M630782	ST73	1	124kb	4	135kb
H309	ST73	3	198kb	6	196kb
H187	ST73	2	147kb	3	142kb
M678433	ST73	4	154kb	5	143kb
H028	ST73	2	116kb	2	152kb
H200	ST73	4	119kb	5	179kb
H486	ST73	2	136kb	4	122kb
H361	ST73	1	151kb	4	151kb
H026	ST73	2	175kb	3	180kb
2-1-P20	ST73	4	156kb	5	162kb
M619205	ST73	3	155kb	5	146kb
M587973	ST73	5	160kb	9	160kb
H437	ST95	2	206kb	3	110kb
H148	ST95	2	120kb	3	114kb

Strain	ST Lineages	Long read results		MobSuite predictions	
		Total number of plasmids	Total plasmid DNA content	Total number of plasmids	Total plasmid DNA content
H104	ST95	1	118kb	2	114kb
M675220	ST95	14	384kb	14	210kb
H648	ST95	1	147kb	3	200kb
H552	ST95	1	148kb	3	143kb
H008	ST95	2	230kb	3	222kb
H219	ST95	1	146kb	2	145kb

Table 4-1 Comparison of Long read sequence results and MobSuite predictions

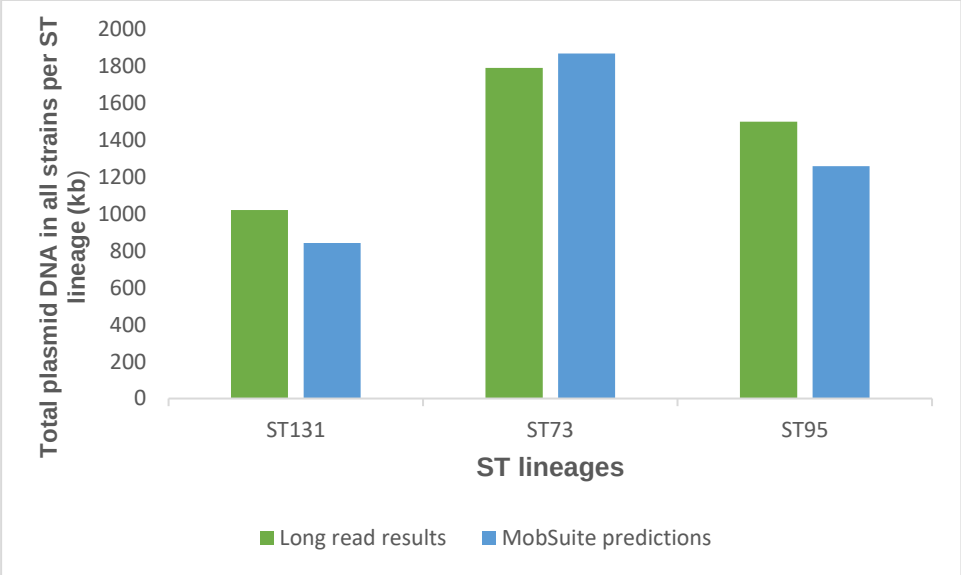

Figure 4.1 Long-read vs MobSuite total plasmid DNA

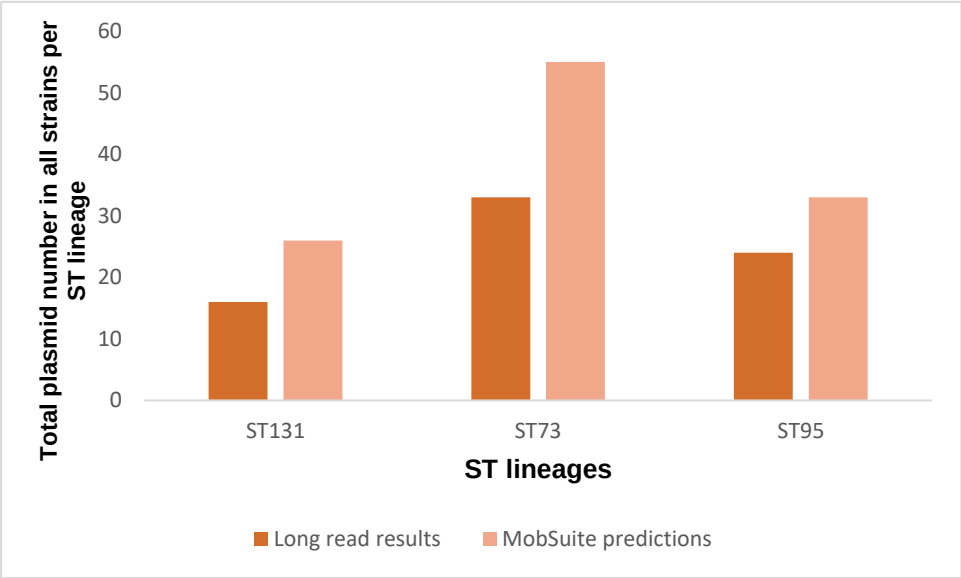

Figure 4.2 Long-read vs MobSuite total number of plasmid

4.4.2 Phylogenetic relationships of the assembled genomes

Twenty-seven (27) assemblies from the long-read data strains were aligned. The strains represent the three ST lineages, with the two common F-plasmid sequence types. F29:A-:B10 is associated with 15 plasmids while F51:A-:B10 is represented by 12 plasmids. The phylogeny shows that same plasmid is circulating in the following gene clusters: H026, H361, M587973, H187, H219, H008, 2-1-P20, M619205, H309, M630782, H648, and H552 with F51:A-:B10, while same plasmid was found in M68551, M653842, M653835, M173503, M678433, M675220, H471, H200, H148, H104, H028, H437 and BS486 with F29:A-:B10. Only M675220 strain from ST95 and BS411 strain from ST131 also with F29:A-:B10 are excluded (fig. 4.3).

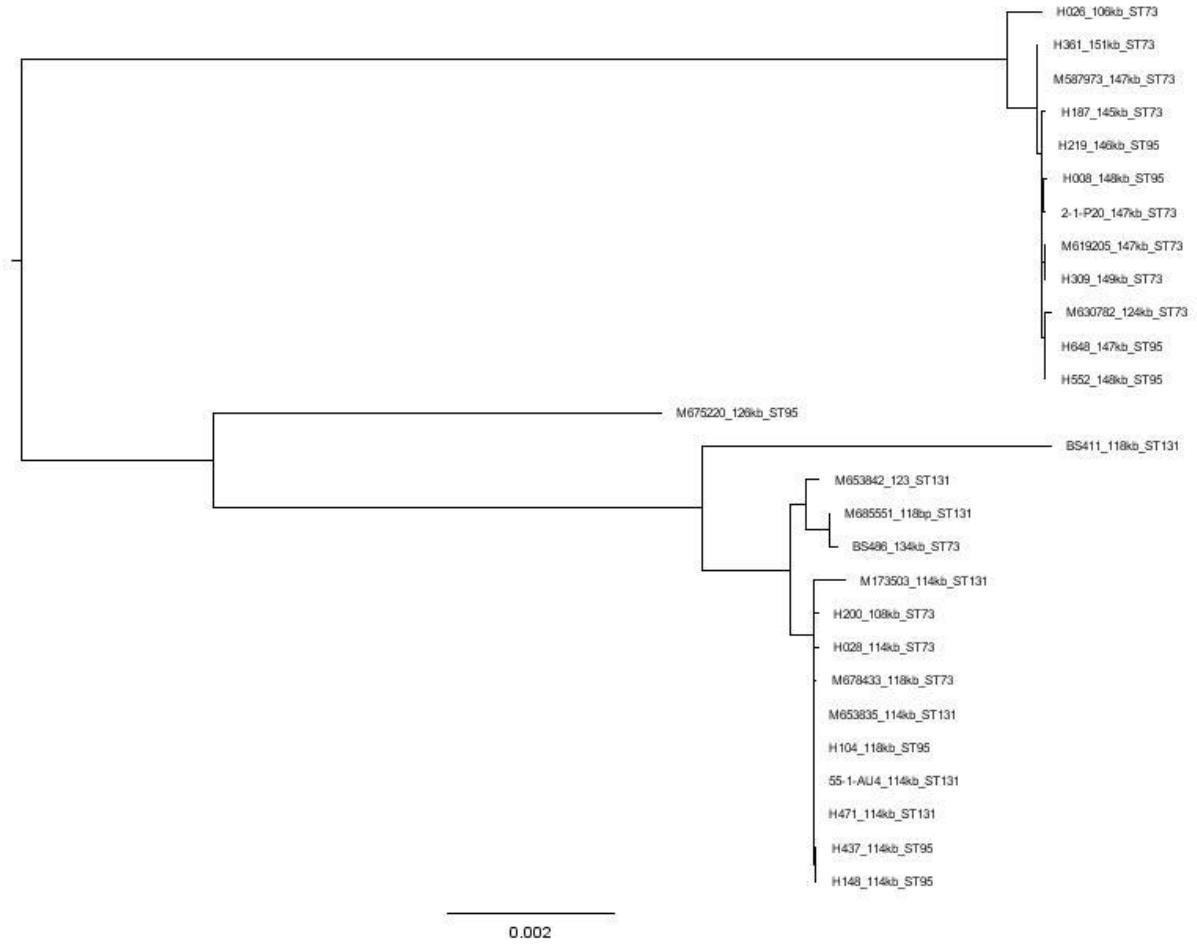

Figure 4.3 **Evolutionary relationships of taxa. Tree is based on the 36 genes that are common to all the plasmids of interest. The tree was constructed using PhyML**

4.4.3 Genomic features associated with selected plasmids

Comparison of the genomic features associated with plasmids of the three lineages were investigated (table 4.2). IncF plasmid was frequently encountered in all genomes especially IncFll and IncFIB. The virulence gene, *traT* is the most common virulence factor detected within the plasmids of the three lineages. Beta-lactam resistance gene, *blaTEM-1B* was also common among the plasmids, although some plasmids do not encode any resistance gene. In summary, the plasmids showed almost the same genomic features even though, they were from different lineages.

Strain	F-plasmid sequence type	ST Lineages	Resistance gene	virulence gene
M685551	F29:A-:B10	ST131	tet(B)	traT, senB
M653842	F29:A-:B10	ST131	blaCTX-M-27	traT, senB
M653835	F29:A-:B10	ST131	Tet(A)	traT, senB
H471	F29:A-:B10	ST131		traT, senB
BS411	F29:A-:B10	ST131	ciprofloxacin	traT, senB
55-1-AU4	F29:A-:B10	ST131		traT, senB
M173503	F29:A-:B10	ST131		traT, senB
M678433	F29:A-:B10	ST73		traT, senB
H028	F29:A-:B10	ST73		traT, senB
H200	F29:A-:B10	ST73		traT, senB
BS486	F29:A-:B10	ST73	blaTEM-1B, sul2, aph(6)-I, aph(3'')-Ib	traT, senB
H437	F29:A-:B10	ST95	blaTEM-1C, sul2, aph(6)-I, aph(3'')-Ib	traT, senB
H148	F29:A-:B10	ST95		traT, senB
H104	F29:A-:B10	ST95		traT, senB
M675220	F29:A-:B10	ST95	blaTEM-30, dfrA17, sul1, sul2, aadA5, aph(6)-I, aph(3'')-Ib	traT, senB
M630782	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB
H309	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB, afaD
H187	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB, afaD
M619205	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB, afaD
M587973	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB, afaD
H361	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB, afaD
H026	F51:A-:B10	ST73	blaTEM-1B	traT, senB
2-1-P20	F51:A-:B10	ST73	blaTEM-1B, sul1, aadA1	traT, senB, afaD
H648	F51:A-:B10	ST95	blaTEM-1B, sul1, aadA1	traT, senB, afaD
H552	F51:A-:B10	ST95	blaTEM-1B, sul1, aadA1	traT, senB, afaD

Strain	F-plasmid sequence type	ST Lineages	Resistance gene	virulence gene
H008	F51:A-:B10	ST95	blaTEM-1B, sul1, aadA1	traT, senB, afaD
H219	F51:A-:B10	ST95	blaTEM-1B, sul1, aadA1	traT, senB, afaD

Table 4-2 Comparison of the genomic features associated with the plasmids in the assembled genomes

4.4.4 Plasmid features of the assembled genomes

The bandage analysis shows the number of plasmids, length, and the plasmid type for each of the assembled genome. The genomes were all shown to consist of at least one circular conjugative plasmid with a range length of between 66kb-151kb. All ST131 strains analysed: H471, M685551, BS411, M653835, M173503, 55-1-TI19 and M653842 have same plasmid, *SCU-122 plasmid pSCU-122-1*. This same plasmid was observed to be circulating in some strains of ST73 (BS486, M678433, H200 and H028), and ST95 (H104, H148 and H437). While some strains of ST73 harbour the plasmid, *PA45B plasmid pPA45B,* strains H187, H309 and M630782 possess similar plasmid, *UPEC 26-1 plasmid unnamed1* with some ST95 (H648, H552, H008, H219) strains.

4.4.5 Plasmid-Host Characteristics

The ST95 and ST73 strains with the same plasmid were found to differ in their serotypes. Serotype O6:H1 and O25:H1 were common in ST73 while serotype O50/O2:H7 was predominant in ST95. They also have different *fimH* types with *FimH41* found only in the strains of ST95 (table 4.3). However, most of the strains in the two lineages share the same sequence type F51:A-:B10. Few strains of ST73 and ST95 share closely related plasmid with the strains of ST131 (fig.4.3). The serotype

and *fimH* type of these strains were however found to differ from that of the strains of ST131 with a predominant serotype, O25:H4 and *FimH30*, *FimH41* and *FimH22*. All ST131 strains, however, share the same F- plasmid sequence type, F29:A-:B10 (table 4.3).

Strain (same plasmid)	**Lineages**	**Serotype**	***FimH* type**	**F-plasmid sequence type**
M630782	ST73	O2:H1	*fimH41*	F51:A-:B10
H309	ST73	O25:H1	*fimH12*	F51:A-:B10
H187	ST73	O6:H1	*fimH30*	F51:A-:B10
M619205	ST73	O25:H1	*fimH12*	F51:A-:B10
M587973	ST73	O6:H1	*fimH10*	F51:A-:B10
H361	ST73	O6:H1	*fimH10*	F51:A-:B10
H026	ST73	O22:H1	*fimH103*	F51:A-:B10
BS486	ST73	O50/O2:H1	*fimH10*	F29:A-:B10
2-1-P20	ST73	O25:H1	*fimH12*	F51:A-:B10
M675220	ST95	O1:H7	*fimH27*	F29:A-:B10
H648	ST95	O50/O2:H7	*fimH41*	F51:A-:B10
H552	ST95	O2:H7	*fimH41*	F51:A-:B10
H008	ST95	O50/O2:H7	*fimH41*	F51:A-:B10
H219	ST95	O50/O2:H7	*fimH41*	F51:A-:B10
Strain (same plasmid)				
M685551	ST131	O25:H4	*fimH41*	F29:A-:B10
M653842	ST131	O25:H4	*fimH30*	F29:A-:B10
M653835	ST131	O25:H4	*fimH22*	F29:A-:B10
H471	ST131	O25:H4	*fimH22*	F29:A-:B10
BS411	ST131	O16:H5	*fimH41*	F29:A-:B10
55-1-AU4	ST131	O25:H4	*fimH22*	F29:A-:B10
M173503	ST131	O25:H4	*fimH30*	F29:A-:B10
H437	ST95	O50/O2:H7	*fimH41*	F29:A-:B10
H148	ST95	O50/O2:H7	*fimH41*	F29:A-:B10
H104	ST95	O18ac:H7	*fimH18*	F29:A-:B10
M678433	ST73	O6:H1	*fimH30*	F29:A-:B10
H028	ST73	O6:H1	*fimH103*	F29:A-:B10

Table 4-3 Comparison of the genetic features of strains with similar ancestry

4.4.6 Mauve analysis of plasmids of F-plasmid sequence types F29:A-:B10 and F51:A-:B10

All plasmids that share similar F-plasmid sequence type displayed highly homologous structures. Comparison of plasmids from the two F-plasmid sequence types also showed some similarities in their structures. AR genes in F51:A-:B10 are very similar. Yet, in F29:A-:B10, these genes are either absent or very diverse. The results further indicate that where there are AR genes, they are not found in the same contig in the same plasmid. However, for similar gene, it was observed that AR genes are always in the same node in the two F-plasmid sequence types. In this analysis, AR gene, *beta-lactamase* was identified in the same contig, in this case, node_2 (fig. 4.6).

Figure 4.4 Plasmids of F29:A-:B10

Figure 4.5 Plasmids of F51:A-:B10

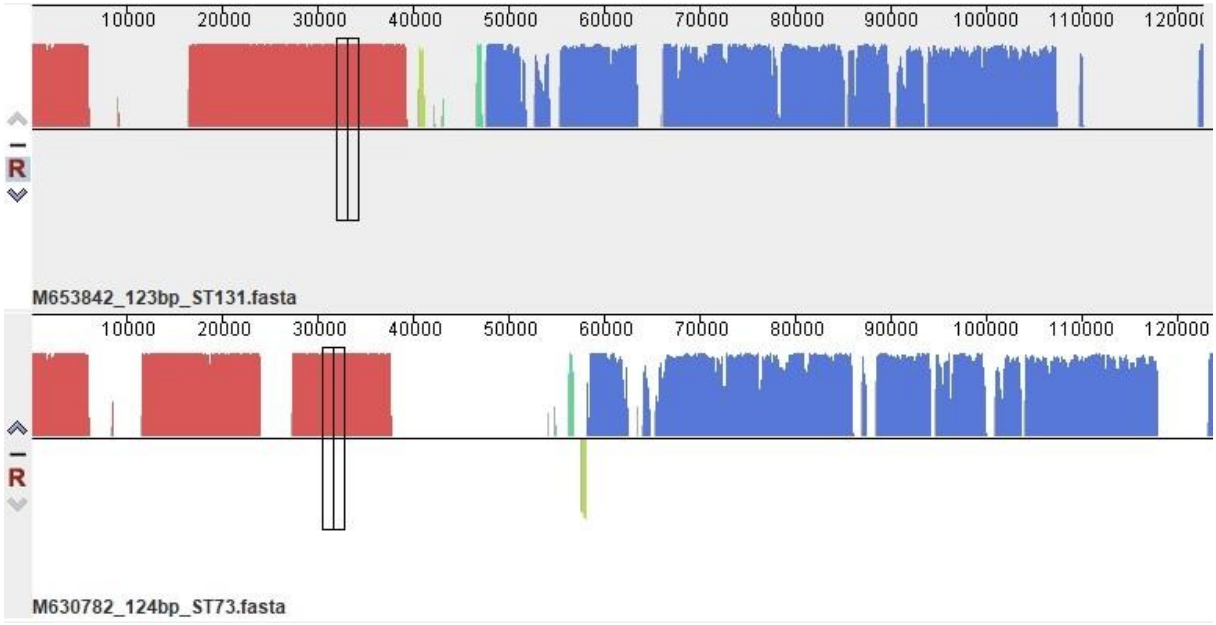

Figure 4.6 Similarities in the plasmids of F29:A-:B10 and F51:A-:B1

4.4.7 Genomic structure comparison

The RAST analysis shows that there are distinct differences in the genome size, number of coding systems and the number of subsystems for the plasmids with the two major F-plasmid sequence types F29:A-:B10 and F51:A-:B10 (table 4.4). The subsystem represents a set of functional roles that are connected to specific genes in specific genomes. The subsystem coverage comprises the breakdown of the number of protein encoding genes with hypothetical and non-hypothetical functions. Common genomic features present in the subsystems of these two key plasmid types are membrane transport that utilises type IV secretion system, RNA metabolism and iron acquisition and metabolism (figs. 4.7 and 4.8). However, most plasmids associated with F51:A-:B10 type have additional specific features: virulence, disease and defence, regulation and cell signalling and presence of integrons (fig. 4.8).

Plasmid	Lineages	F-plasmid sequence type	Size	Number of subsystems	Genomic DNA base Composition	Number of Coding Sequences
55-1-AU4	ST131	F29:A-:B10	114kb	4	51	160
M173503	ST131	F29:A-:B10	114kb	4	51	163
BS411	ST131	F29:A-:B10	118kb	4	50.9	173
H471	ST131	F29:A-:B10	114kb	4	51	159
M653835	ST131	F29:A-:B10	114kb	4	51	163
M653842	ST131	F29:A-:B10	123kb	5	51.1	178
M685551	ST131	F29:A-:B10	118kb	5	50.6	176
M678433	ST73	F29:A-:B10	118kb	4	50.9	158
BS486	ST73	F29:A-:B10	134kb	7	52.1	197
H028	ST73	F29:A-:B10	114kb	4	51	157
H200	ST73	F29:A-:B10	110kb	4	51	155
H104	ST95	F29:A-:B10	118kb	4	51	164
H148	ST95	F29:A-:B10	114kb	4	51	157
M675220	ST95	F29:A-:B10	125kb	8	52.4	179
H437	ST95	F29:A-:B10	114kb	4	51	159
H008	ST95	F51:A-:B10	148kb	9	51.5	207
H552	ST95	F51:A-:B10	148kb	9	51.5	206
H648	ST95	F51:A-:B10	147kb	9	51.6	210
H361	ST73	F51:A-:B10	151kb	10	51.7	221
H026	ST73	F51:A-:B10	106kb	4	51.4	159
H187	ST73	F51:A-:B10	145kb	8	51.5	203
H309	ST73	F51:A-:B10	149kb	9	51.5	205
M587973	ST73	F51:A-:B10	147kb	9	51.6	210
M630782	ST73	F51:A-:B10	124kb	9	51.9	172
M619205	ST73	F51:A-:B10	147kb	9	51.6	209
2-1-P20	ST73	F51:A-:B10	148kb	9	51.6	212

Table 4-4 RAST analysis showing genomic attributes of the plasmids of the ST lineages

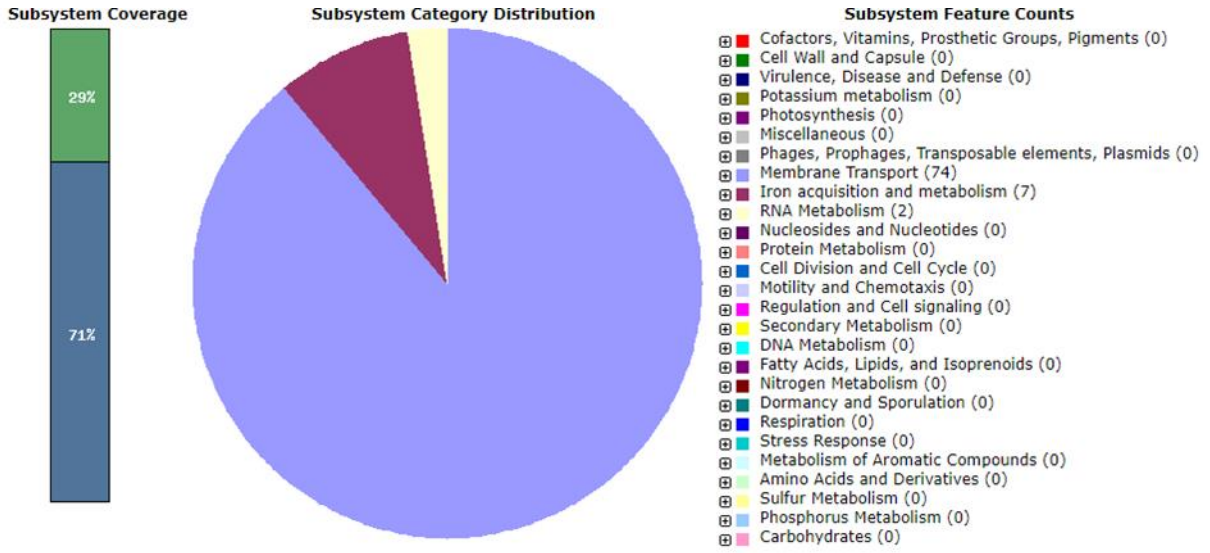

Figure 4.7 **Genes connected to subsystems and their distribution in different categories for most F29:A-:B10 plasmid**

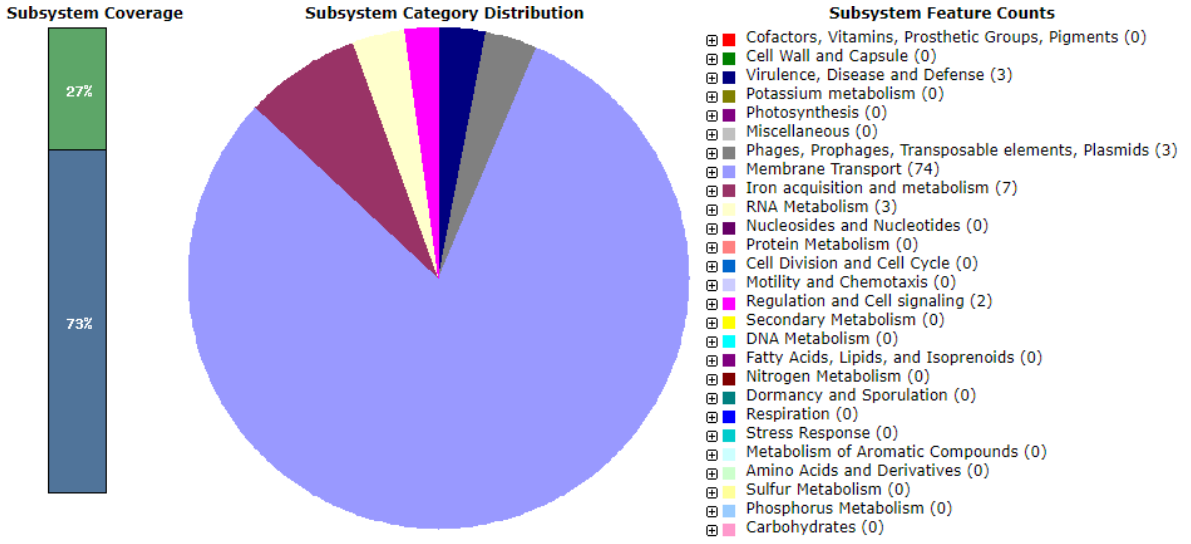

Figure 4.8 **Genes connected to subsystems and their distribution in different categories for most F51:A-:B10 plasmid**

4.4.8 Distribution of genes within the key F-plasmid sequence types

The Roary analysis reveals that within the pangenome of these plasmids totaled 366 genes, only 36 genes were common to all the plasmids (Table 4.5).

In addition, the analysis shows that the F-plasmid sequence types encode multiple genes: multi-functional conjugation genes *traI*, *traF*, pilin gene *traA*, protein coupling genes *traD*, *traC*, *traQ* and *traN* as well as *repB* were all present at almost the same proportion in the two major F-plasmid sequence types F29:A-:B10 and F51:A-:B10. There were significant differences in the proportion of *traV* and *traJ* genes, relaxosome genes, *traM* and *traY*, fertility inhibition gene *finO*, toxic gene *hokB,* antitoxin *ccdA* and toxin *ccdB* genes found in F29:A-:B10 (92.86% for *traV, traJ, traM*, 85.71% for *traY* and 100% for *finO*, *hokB*, *ccdA* and *ccdB* genes) when compared with F51:A-:B10 (7.69% for all the genes). There were also significant differences in the multi-drug transporter gene *emrE*, endoribonuclease toxin *chpB* and antitoxin *pemI* detected in the two plasmid types. In addition, multi-drug resistance gene *mdtH*, virulence regulon transcriptional activator genes, *virF* and *virB*, transposase *InsE-1_1*, plasmid partition gene A, *parA*, and afimbrial adhesin gene, *afaD* were found only in the F-plasmid sequence type F51:A-:B10 at significant proportions (table 4.6).

Pangenome	**Percentage occurrence**	**Number of genes**
Core genes	99% <= strains <= 100%	36
Soft core genes	95% <= strains < 99%	6
Shell genes	15% <= strains < 95%	204
Cloud genes	0% <= strains < 15%)	120

Table 4-5 Gene distribution within the genomes of the STs

Gene	**F-plasmid sequence type**	
	F29:A-:B10	**F51:A-:B10**
traI	100%	100%

Gene	F-plasmid sequence type	
	F29:A-:B10	**F51:A-:B10**
traQ	100%	100%
traC	100%	100%
traN	100%	100%
traA	100%	100%
traT	100%	100%
traR	100%	100%
traF	100%	100%
traH	100%	100%
traD	86%	100%
traV	93%	8%
traJ	93%	8%
traM	93%	8%
traY	86%	8%
traL	7%	0%
traG	7%	0%
repB	100%	100%
ylpA	100%	100%
nqrC	93%	92%
peml	29%	85%
finO	100%	8%
hokB	100%	8%
umuD	100%	8%
parA	0%	92%
parM	100%	8%
ccdA	100%	8%
ccdB	100%	8%
chpB	29%	85%
mdtH	0%	85%
emrE	7%	85%
afaD	0%	69%
virB	0%	92%
virF	0%	77%
insE-1_1	0%	85%

Table 4-6 Gene content comparisons of the major F-plasmid sequence types

4.5 Discussion

In this study, we summarise long-read assembly data of the genomes of the phylogroup B2 lineages based on how many plasmids were in a strain, plasmid DNA content, F-plasmid sequence type, type of plasmid found in each strain, and similarities and differences between the long-read sequence data results and MobSuite predictions. In addition, the characteristics of the two common plasmids, and their phylogenetic relationships were defined and compared. The two F-plasmid sequence types: F29:A-:B10 and F51:A-:B10 shared by most of the plasmids of the three lineages were used for this analysis.

Differences were observed in the total plasmid DNA and number of plasmids in the assembled genomes due to the type of genome sequencing method used to analyse them. Although long-read results reveal that about 33% of all assemblies harbour a single conjugative plasmid, plasmid number and DNA content differs: long-read data results show that the number of plasmids found across the strains were between 1-14, with a total plasmid DNA content range of between 114kb – 384kb. In contrast, results from MobSuite predictions show plasmids of between 2-14, and about 109kb – 222kb plasmid DNA content. MobSuite predictions tend to be more accurate in estimating plasmid DNA contents and plasmid number.

In the long-read analysis, plasmids associated with the two common F-plasmid sequence types display different characteristics. 33% of ST131, 12% of ST73 and 17% of ST95 strains harboured F29:A-:B10 plasmid while 37% of ST73 and 20% of ST95 strains harboured F51:A-:B10 plasmid (chapter 2). F51:A-:B10 plasmids are quite homogeneous unlike the F29:A-:B10 plasmids. The average plasmid DNA content of

most conjugative plasmids of F51:A-:B10 is 142kb. These plasmids also have varying similar resistance genes. β-lactamase, *blatem-1B*, sulphonamide*, sul1* are prevalent in this sequence type. These genes were found encoded in IncFII and IncFIB replicons, common in all plasmids analysed. This replicon combination has been previously reported as a multi-replicon, often observed on a single plasmid (Pierre-Emmanuel *et al*., 2020; Villa *et al*., 2010). In F29:A-:B10 plasmids, the average plasmid DNA content of most of the conjugative plasmids is 117kb, with resistance genes either absent or very diverse. These characteristics play a key function in the evolution of these fertility plasmid types.

Besides enhancing conjugal transfers in the lineages (Lan *et al*., 2001), *tra* genes encoded in these plasmid types ensure the propagation and maintenance of the plasmids in the bacterial populations. Since the persistence of plasmids within the host is dependent on host fitness (Khosravi *et al*., 1990), the many advantageous effects of the *tra* genes encoded in the plasmids can enhance the adaptation and evolution of the bacterial host. The *traA* gene encodes propilin, utilised for the formation of conjugative pili, needed for cell-to-cell contact during the initial stages of bacterial conjugation (Paiva *et al*., 1992). The gene, *traT* is a key F-factor transfer gene. It harbours an outer membrane protein used by *E. coli* to exert an influence in its environment (Achtman *et al*., 1977). Propilin, encoded by *traA* interacts with *traQ* to assemble mature pilin (Kathir and Ippen-Ihler, 1991). The assembly of mature F-pilin subunits into extended F-pili requires the role of *traC* (Schandel *et al*., 1990). The initial process of transfer DNA synthesis at the origin of transfer is promoted by *traI* (Ziegelin *et al*., 1991). Conjugal DNA transfer requires the help of *traD* and *traN* (Sastre *et al*, 1998; Pohlman *et al*., 1994), while *repB* is essential for plasmid replication (Gibbs *et*

al., 1993). These genes are proportionally distributed in the two F-plasmid sequence types.

Diversities were also observed in the gene contents of the two F-plasmid sequence types. *traJ*, *traM* and *traY* genes were not common in plasmids with F51:A-:B10. The two components of the relaxosome, *traY* and *traM* with *traJ*, ensure interactions with *traI* genes for site- and -strand-specific cut at the origin of transfer during bacterial conjugation (Penfold *et al*., 1996; Nelson *et al*., 1995). Regions in F-plasmid oriT locus are bound by *traM.* These regions facilitate plasmid transfer and auto-regulation (Di Laurenzio *et al*., 1992). Relaxosome formation is aided by *traY* by binding to the origin of transfer region. It also helps to control *tra* gene expression (Luo *et al*., 1994). Other genes not prevalent in F51:A-:B10 are *traV* which is essential for F pilus assembly (Skyberg *et al*., 2006), *finO* gene, which inhibits the expression of *traJ* (Ghetu *et al.,* 2000), required to regulate the expression of transfer genes that are needed in the conjugal transfer of DNA between bacterial cells (Zahrl *et al*., 2007), *hokB*, a toxic component of a type I toxin-antitoxin system that can eliminate bacterial cells within minutes, once overexpressed (Pedersen and Gerdes, 1999), *parM which* helps in the regulation of plasmid partition, and is essential for accurate plasmid segregation (van den Ent *et al*., 2002), *ccdB*, a toxic component of a type II toxin-antitoxin system that plays a major role in plasmid maintenance, especially in post-segregational killing of plasmid-free cells (Bernard and Couturier, 1992) and *ccdA*, the antitoxin component of type II toxin-antitoxin system that prevents post-segregational killing of plasmid-free cells, by blocking the activity of *ccdB (*Maki *et al.,* 1996). These genes were all present in higher proportions in plasmids with F29:A-:B10. On the other hand, while some of the genes were not detected in any plasmid borne F29:A-:B10 sequence type, some

genes were over-represented in F51:A-:B10 encoded plasmids. Genes such as *mdtH* that confers resistance to norfloxacin and enoxacin (Nishino and Yamaguchi, 2001), *virB* and *virF,* which are regulators of plasmid encoded virulence genes (Tobe *et al*., 1993), *insE-1_1*, involved in transposition of the insertion sequence IS3 (Tobe *et al*., 2020), *parA,* crucial for plasmid partition, and enables the effective distribution of newly replicated plasmids into daughter cells during cell division (Abeles *et al*, 1985), afimbrial adhesin gene (*afaD*) that enables the production of afimbrial adhesive sheath. This gene encodes adhesin–invasins, needed for adherence to epithelial cells for cytolysis and defence responses (Pichon *et al*., 2013; Garcia *et al*., 2000), *emrE,* a multidrug efflux gene that confers resistance to a wide range of toxic compounds (Ninio *et al*., 2001; Yerushalmi and Schuldiner, 2000), *ChpB,* the toxic component of a type II toxin-antitoxin system whose endoribonuclease activity is inhibited by the antitoxin *chpS* (Christensen *et al.,* 2003; Masuda *et al*., 1993) and *pemI,* an antitoxin component of a type II toxin-antitoxin system that binds to the toxin, *pemK*, and neutralizes its activity, and also helps in plasmid maintenance during cell division (Tsuchimoto and Ohtsubo, 1993; Tsuchimoto *et al*., 1988).

Toxin-antitoxin genes ensure the growth of plasmid-carrying daughter cells in a bacterial population by killing new cells that have not inherited a plasmid copy at cell division. These genes were more prevalent in F29:A-:B10 plasmids. The system enables the stabilization and maintenance of the plasmids associated with the F-plasmid sequence types (Jensen and Gerdes, 1995) and also promote the stabilization of the host cell by inhibiting harmful fatal deletions in the cell *(*Szekeres *et al.,* 2007*;* Rowe-Magnus *et al*., 2003). These functions enhance the pathogenic and pandemic nature of this F-plasmid sequence as well as improve the evolution of the bacterial

host. However, in most of the cells that harbour two or more plasmids from the same incompatibility group, with each daughter cell carrying a plasmid, often, when one of these plasmids encode for a TA system, and it is displaced by a TA-free plasmid system, the plasmid becomes unstably inherited, and thus post-segregational killing is induced (Ogura and Hiraga, 1983). This process may have possibly led to the instability of some of the F29:A-:B10 plasmids. But the cells were still able to survive due to the presence of specific toxin- antitoxin systems, *ccdA* and *ccdB*, that were not found in F51:A-:B10 plasmids. The *ccdA* antitoxin can allow the cell to lose a plasmid without experiencing the effects of the toxin it encodes. The antitoxin immobilizes the *ccdB* toxin, thereby, preventing the activation of the toxin when the plasmid is lost (Ramisetty, and Santhosh, 2016; Bahassi *et al*., 1999; Bernard and Couturier, 1992; Ogura and Hiraga, 1983). Overall, our study shows that F29:A-:B10 and F51:A-:B10 plasmids differ in their genomic features, but genes encoded within these pMLST types, play numerous unique roles in conjugal gene transfer, plasmid stability and host fitness (chapter 3).

4.6 Conclusion

Long-read and short-read sequencing are essential tools in phylogenetic studies, but data from long-read sequencing are crucial when investigating evolutionary relationship existing among different lineages. This study shows that the two F-plasmid sequence types common in the strains of the three ST lineages differ in their plasmid-host characteristics and gene contents. The diversities observed can be explained by the host genetic background and differences in encoded mobile genetic elements.

Despite the differences, the encoded genes play multifunctional roles in the plasmid maintenance and host cell adaptation and evolution.

4.7 References

Abeles, A. L., Friedman, S. A. and Austin, S. J. (1985). "Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region." *J Mol Biol* **185**, 261-272.

Achtman, M., Kennedy, N. and Skurray, R. (1977). Cell-Cell interactions in conjugating *Escherichia coli*: Role of *TraT* protein in surface exclusion. *Proc Natl Acad Sci* **74**, 5104–5108.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) "Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics.*

Bernard, P. and Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes." *J Mol Biol* **226**, 735-745.

Carattoli, A., Zankari, E., Garcia-Fernandez, A., Voldby- Larsen, M., Lund, O., Villa, L., Aarestrup, F. M. and Hasman, H. (2014). PlasmidFinder and pMLST: in silico detection and typing of plasmids. *Antimicrob. Agents Chemother* **10**, 1128.

Christensen, S. K., Pedersen, K., Hansen, F .G. and Gerdes, K. (2003). "Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA." *J Mol Biol* **332**, 809-819.

Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. (2004). Mauve: multiple alignment Of conserved genomic sequence with rearrangements. *Genome research* **14**, 1394-1403.

Darling, A. E., Mau, B., and Perna, N. T. (2010). ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **25**, 5e11147.

Di Laurenzio, L., Frost, L. S. and Paranchych, W. (1992). "The *TraM* protein of the conjugative plasmid F binds to the origin of transfer of the F and ColE1 plasmids. *Mol Microbiol* **6**, 2951- 2959.

Garcia, M. I., Jouve, M., Nataro, J., Gounon, P. and Bouguénec, C. (2000). Characterization of the AfaD-like family of invasins encoded by pathogenic *Escherichia coli*. *FEBS letters* **479**, 111-7.

Ghetu, A. F., Gubbins, M. J., Frost, L. S. and Glover, J. N. (2000). "Crystal structure of the bacterial conjugation repressor *finO*." *Nat Struct Biol* **7**, 565-569.

Gibbs, M. D., Spiers, A. J. and Bergquist, P. L. (1993). "RepFIB: a basic replicon of large plasmids." *Plasmid* **29**, 165-179.

Guindon, S. and Gascuel, O. (2003). A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704.

Jensen, R. B., and Gerdes, K. (1995). Programmed cell death in bacteria: Proteic plasmid stabilization systems. *Mol Microbiol* **17**, 205-210.

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M. and Aarestrup, F. M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Micobiol* **52**, 1501-1510.

Johansson, M. H. K., Bortolaia, V., Tansirichaiya, S., Aarestrup, F. M., Roberts, A. P. and Petersen, T. N. (2021). Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J Antimicrob Chemother* **76**, 101-109.

Kathir, P. and Ippen-Ihler, K. (1991). Construction and characterization of derivatives carrying insertion mutations in F plasmid transfer region genes, *trbA, artA, traQ*, and *trbB*. *Plasmid* **26,** 40–54.

Khosravi, M., Ryan, W., Webster, D. A. and Stark, B. A. (1990). Variation of Oxygen Requirement with Plasmid Size in Recombinant *Escherichia coli*. *Plasmid* **23,** 138-143.

Lan, R., Lumb, B., Ryan, D. and Reeves, P. (2001). Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli Infect Immun* **69**, 6303-6309.

Luo, Y., Gao, Q. and Deonier, R. C. (1994). "Mutational and physical analysis of F plasmid *traY* protein binding to oriT." *Mol Microbiol* **11***,* 459-469.

Maki, S., Takiguchi, S., Horiuchi, T., Sekimizu, K. and Miki, T. (1996). "Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (*CcdB*) and LetA (*CcdA*)." *J Mol Biol* **256**, 473-482.

Masuda, Y., Miyakawa, K., Nishimura, Y. and Ohtsubo, E. (1993). chpA and chpB, *Escherichia coli* chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100." *J Bacteriol* **175**, 6850-6856.

Morrison, D. A. (2006). Multiple sequence alignment for phylogenetic purposes. *Austral Syst Bot* **19**, 479-539.

Morrison, D. A. (2011). Phylogenetic Analysis of Pathogens. Editor(s) Michel Tibayrenc Genetics and Evolution of Infectious Disease. *Elsevier* 203-231. ISBN 9780123848901.

Nelson, W. C., Howard, M. T., Sherman, J. A. and Matson, S. W. (1995). "The *traY* gene product and integration host factor stimulate *Escherichia coli* DNA helicase I-catalyzed nicking at the F plasmid oriT." *J Biol Chem* **270**, 28374-28380.

Nicolas-Chanoine, M. H., Blanco, J., Leflon-Guibout, V., Demarty, R., Alonso, M. P., Caniça, M. M., Park, Y. Y., Lavigne, J. P., Pitout, J. and Johnson, J. R. (2008). Intercontinental Emergence of *Escherichia coli* Clone O25:H4-ST131 Producing CTX-M-15. *J Antimicrob Chemother* **61**, 273-281.

Ninio, S., Rotem, D. and Schuldiner, S. (2001). "Functional analysis of novel multidrug transporters from human pathogens." *J Biol Chem* **276**, 48250-48256.

Nishino, K. and Yamaguchi, A. (2001). "Analysis of a complete library of putative drug transporter genes in *Escherichia coli*." *J Bacteriol* **183**, 5803-5812.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**, 206-214.

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**, 3691–3.

Paiva, W. D., Grossman, T. and Silverman, P. M. (1992). "Characterization of F-pilin as an inner membrane component of Escherichia coli K12." *J Biol Chem* **267**, 26191-26197.

Pedersen, K. and Gerdes, K. (1999). Multiple hok genes on the chromosome of *Escherichia coli." Mol Microbiol* **32***,* 1090-1102*.*

Penfold, S. S., Simon, J. and Frost, L. S. (1996). "Regulation of the expression of the *traM* gene of the F sex factor of *Escherichia coli*." *Mol Microbiol* **20**, 549-558.

Pichon, C., du Merle, L., Lequeutre, I. and Le Bouguénec, C. (2013). The AfaR small RNA controls expression of the AfaD-VIII invasin in pathogenic *Escherichia coli* strains. *Nucleic acids research* 41, 5469–5482.

Pierre-Emmanuel, D., Ludovic, M., Nicolas, R., Arnaud, F. and Michel-Yves, M. (2020). Analysis of COMPASS, a New Comprehensive Plasmid Database

Revealed Prevalence of Multi-replicon and Extensive Diversity of IncF Plasmids. *Frontiers in Microbiology* **11,** 483.

Pohlman, R. F., Genetti, H. D. and Winans, S. C. (1994). "Common ancestry between IncN conjugal transfer genes and macromolecular export systems of plant and animal pathogens." *Mol Microbiol* **14**, 655-668.

Rowe-Magnus, D. A., Guerout, A. M., Biskri, L., Bouige, P. and Mazel, D. (2003)."Comparative analysis of superintegrons: engineering extensive genetic diversity in the *Vibrionaceae.*" *Genome Research* **13**, 428–42.

Sastre, J. I., Cabezon, E., de la Cruz, F. (1998). “The carboxyl terminus of protein *TraD* adds specificity and efficiency to F-plasmid conjugative transfer." *J Bacteriol* **180**, 6039-6042.

Schandel, K. A., Maneewannakul, S., Vonder Haar, R. A., Ippen-Ihler, K. and Webster, R. E. (1990). Nucleotide sequence of the F plasmid gene, *traC*, and identification of its product." *Gene* **96**, 137-140.

Skyberg, J. A., Johnson, T. J., Johnson, J. R., Clabots, C., Logue, C. M. and Nolan, L. K. (2006). "Acquisition of avian pathogenic Escherichia coli plasmids by a commensal E. coli isolate enhances its abilities to kill chicken embryos, grow in human urine and colonize the murine kidney." *Infect Immun* **74**, 6287-6292.

Szekeres, S., Dauti, M., Wilde, C., Mazel, D. and Rowe-Magnus, D. A. (2007). "Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection". *Molecular Microbiology* **63**, 1588–605.

Tobe, T., Hayashi, T., Han, C. G., Schoolnik, G. K., Ohtsubo, E. and Sasakawa, C. (2020). "Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid." *Infect Immun* **67**, 5455-5462.

Tobe, T., Yoshikawa, M., Mizuno, T. and Sasakawa, C. (1993). Transcriptional control of the invasion regulatory gene virB of *Shigella flexneri*: activation by virF and repression by H-NS." *J Bacteriol* **175**, 6142-6149.

Tsuchimoto, S. and Ohtsubo, E. (1993). "Autoregulation by cooperative binding of the PemI and PemK proteins to the promoter region of the pem operon." *Mol Gen Genet* **237**, 81-88.

Tsuchimoto, S., Ohtsubo, H. and Ohtsubo, E. (1988). Two genes, pemK and pemI, responsible for stable maintenance of resistance plasmid R100." *J Bacteriol* **170**, 1461-1466.

Van den Ent, F., Moller-Jensen, J., Amos, L. A., Gerdes, K. and Lowe, J. (2002). "F-actin-like filaments formed by plasmid segregation protein *ParM*." *EMBO J* **21**, 6935-6943.

Villa, L., García-Fernández, A., Fortini, D. and Carattoli, A. (2010). Replicon Sequence Typing of IncF Plasmids Carrying Virulence and Resistance Determinants. *Journal of Antimicrobial Chemotherapy* **65**, 2518–2529.

Wick, R. R., Schultz, M. B., Zobel, J. and Holt, K. E. (2015). “Bandage: interactive visualization of de novo genome assemblies. “*Bioinformatics* **31**, 3350-2.

Yerushalmi, H. and Schuldiner, S. (1995). "An essential glutamyl residue in *EmrE*, a multidrug antiporter from *Escherichia coli*." J. Biol Chem 270, 6856-6863.

Zahrl, D., Wagner, A., Tscherner, M. and Koraimann, G. (2007). “GroEL plays a central role in stress-induced negative regulation of bacterial conjugation by promoting proteolytic degradation of the activator protein *TraJ*." *J Bacteriol* **189**, 5885-5894.

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. *J Antimicrob Chemother* **67**, 2640–2644

Ziegelin G, Pansegrau W, Strack B, Balzer, D., Kroger, M., Kruft, V. and Lanka, E. (1991). Nucleotide sequence and organization of genes flanking the transfer origin of promiscuous plasmid RP4. DNA Sequence : *J DNA Sequencing and Mapping* **1**, 303-327.

Chapter 5 General Discussion

5.1 Antibiotics susceptibility in some human associated *E. coli* strains

This research highlighted the reasons why strains of ST73 and ST95 lack antibiotics resistance. Most plasmids of ST73 are small mobilisable plasmids, which in most cases are about 40kb or less. Such small plasmid DNAs lack integrons required for the capture and expression of resistance genes. In ST73, 29% of these plasmids do not belong to any IncF group and only 55% harbour conjugative systems. Both conjugation systems and IncF groups are required for genetic transfers during conjugation. In addition, fluoroquinolone (FQ) resistance is generally associated with changes in the chromosomal genes *gyrB* and *parC*. The lack of FQ resistance in ST73 and ST95 is one reason these STs are less likely to be multidrug resistant. Furthermore, *dfrA17* and *aadA5* are genes cassettes not common in ST73 and ST95. These gene cassettes are both known to confer resistance. Most strains are characterized majorly by the F-plasmid sequence types, F51:A-:B10 and F29:A-:B10. These plasmids can be pathogenic, they can carry genes that enhance virulence.

5.2 ExPEC Strains and antibiotic resistance

Escherichia coli is a common bacterium in the family Enterobacteriaceae. Although, strains *of E. coli* can be harmless commensals inhabiting the intestines of humans and animals (Kaper *et al.*, 2004), the extra-intestinal pathogenic *E. coli* (EXPEC) strains can become opportunistic pathogens when they are able to get into other parts of the body where they do not belong, causing infections in these extra-intestinal sites:

bloodstream, bladder, kidneys, brain, spinal cord, and the lungs. These pathogenic strains have been implicated in a vast number of community and hospital acquired infections globally (Skov and Monnet, 2016; Liu *et al*., 2016; Nicolas-Chanoine *et al*., 2014; Chen *et al*., 2014; Golding *et al*., 2012; Dias *et al*., 2009; Coque *et al*., 2008; Nicolas-Chanoine *et al*., 2008). ExPEC strains are responsible for a wide range of highly life-threatening diseases: urinary tract infections, neonatal meningitis, pneumonia, and bacteraemia (WHO, 2020; Jang *et al.*, 2017; Fratamico *et al*., 2016).

Records from World Health Organization has shown that the global increase and proliferation of antimicrobial resistance (AMR), especially the resistance of antibiotics that were usually used to treat Enterobacteriaceae causing diseases have become a great concern to world economic growth and health High rates of resistance to Fluoroquinolone and Ciprofloxacin antibiotics used to treat UTIs caused by *E. coli and Klebsiella pneumonia* have been globally recorded. Colistin, an antibiotic used to treat carbapenem resistant Enterobacteriaceae, has also become resistance in many regions. In addition, reports of high rates of resistance of *E. coli* to third generation cephalosporins, used to treat bloodstream infections have also been documented (AMR; WHO, 2020). Coupled with this, is the adaptation and evolution of Enterobacteriaceae bacteria to multi-drug resistance (MDR) (AMR; WHO, 2020).

Phylogroup B2 is the sub-group of *E. coli* strains responsible for most pathogenic *E. coli* diseases. This phylogroup typically consists of three major human associated sequence types (STs) lineages: ST131, ST95 and ST73. Many reports have demonstrated that each of these three lineages play equal role in ExPEC infections, with each representing about 10% of human associated *E. coli* diseases (Forde *et al*.,

2019; Gordon *et al*., 2017; Kallonen *et al*., 2017; Stephens *et al*., 2017; Chen *et al*., 2014;Gordon, 2010; Nicolas-Chanoine *et al*., 2008). Given how ubiquitous these strains are, it is anticipated that they all should be exposed to antimicrobials at about the same rate. However, numerous studies have shown that despite their globally pathogenic nature and their capacity to equally cause extra intestinal diseases, there are variations in their antimicrobial profiles : strains of ST131 are highly multi- drug resistance (Kallonen *et al.*, 2017; Johnson *et al*., 2016), while the other two lineages, STs 73 and 95 are largely antibiotic susceptible (Kallonen *et al*., 2017; Stephens *et al*., 2017; Gordon *et al*., 2017). The reasons for the antibiotic susceptibility of these ExPEC *E. coli* strains remains largely unknown and were therefore investigated.

5.3 Plasmid diversity within *E. coli* lineages

The genes conferring resistance to antimicrobials are often plasmid mediated. F-plasmids also known as fertility plasmids are often responsible for antibiotic resistance. These plasmid types are common in the Enterobacteriaceae family. In this study (chapter 2), we found that plasmids were well distributed in the three lineages with at least three plasmids in most strains of each lineage. These plasmids were observed as small mobilisable plasmids in the strains of ST73 lineage. Some strains of this lineage have plasmid DNA content of less than 40kb. These plasmids cannot self-transfer during conjugation, unlike the large conjugative plasmids in most strains of ST131 and some strains of ST95. These conjugative plasmids, IncFII and IncFIB, although also observed in some strains of ST73, were not common. Two common F-plasmid sequence types were shared among the three lineages. Most plasmids with F51:A-:B10 sequence type encode numerous integrons, but few antimicrobial

resistance (AMR) genes. Integrons are important genetic elements in gene acquisition and expression (Mazel, 2006). Although integrons were not common in F29:A-:B10 plasmids, the plasmids of ST131 that encode integrons possess large number of AMR genes. O25H4:*fimH30*, O16H5:*fimH41* are prevalent characteristics of most strains of ST131.These serotype and *fimH* types are adapted to integrons and multiple resistance genes including a fluoroquinolone resistance determinant, not observed in the plasmids of STs 73 and 95. Plasmids encoded by ST73 and ST95 were more associated with *blaTEM-1* B, *sul1* and *Aada1* genes. In addition to integrons and AMR genes, plasmids of the three lineages also possess many plasmid encoded virulence genes and bacteriocins. Most of these virulence determinants are involved in bacterial conjugation and regulation**,** enterotoxin production, cytotoxic necrotizing, biosynthesis of aerobactin, cell adhesion, proteolysis, iron acquisition and regulation *Colicin E1* and *Colicin V* were also common in some plasmids of all the three lineages, while *colicin ia, ib* and *microcin B17* were observed only in ST95 and ST131 plasmids. The bacteriocins are responsible for cytolysis and defence response of cell to adverse conditions. Overall, unique diversity of plasmid characteristics are harboured in the three lineages.

5.4 Plasmids and conjugal transfers

Plasmids are essential for bacteria evolution and adaptation (Lan *et al*., 2001). They play very important roles in conjugal gene transfer (Sherley *et al*., 2003). These plasmids allow genes to be transferred from donor bacterium, carrying the F-factor to the recipient bacterium, devoid of the F-factor by conjugal transfer *(*Johnston *et al*., 2014; Stearns and Hoekstra, 2005; Lawley *et al.,* 2003*;* Alberts *et al*., 2002). This study (chapter 3) demonstrated that the host genomic background was not influencing the

rates of plasmid transfer rather, small differences observed in the plasmid transfer rate in relation to F-plasmid sequence types, were as a result of the presence of multiple *tra* genes harboured by most plasmids of F29:A-:B10 and F51:A-:B10. Conjugal transfer is an important means of gene spread. The *tra* genes associated with the F-plasmid sequence types were transferred in the conjugation process ensuring the dissemination and evolution of the genes. Therefore, the distinct functions of these *tra* genes in conjugal plasmid transfer is pivotal to the *E. coli* lineages adaptation and evolution.

5.5 Plasmid stability and host fitness

Although plasmids confers unique benefits to bacterial host, the presence of plasmids within a bacterial host can often be a burden to the host cell (Baltrus, 2013; Shachra *et al*., 2010 and Doyle *et al*., 2007). How well plasmids are maintained in the host will thus depend among other things, on the host fitness. When the costs of carrying the plasmid exceeds its usefulness and there is negative selection, it is expected that plasmid free cells will out-compete plasmid-bearing cells (Warnes and Stephenson,1986; Lee and Edlin,1985). Although some ST131 plasmids were unstably inherited, differences were observed in the plasmid stability of the three ST lineages. These differences were also observed in plasmid stability in relation to the two common F-plasmid sequence types. The plasmid acquisition cost, which possibly arose because of the presence of the multiple *tra* genes as well as the host cell characteristics may have affected the stability of the plasmids especially ST131 plasmids that was noted to encode higher number of resistance genes.

5.6 Genomes of the two common F-plasmid sequence types

The findings from this study (chapters 2 and 3) have shown that numerous distinct plasmids encoding varying genetic elements are present among the three lineages of phylogroup B2, and these plasmids are mostly associated with two key F-plasmid sequence types. In the final analysis (chapter 4), we investigated and compared the evolutionary history of the two common F-plasmid sequence types in the plasmids of the three lineages using genome alignments techniques and phylogenetic analysis. Similar clusters of plasmids were associated with either F29:A-:B10 or F51:A-:B10 plasmids. None of the ST131 genomes harbour the F51:A-:B10 plasmid (chapter 2), and only ST73 and ST95 with F29:A-:B10, share similar plasmid with ST131 strains.

Significant differences were observed in the gene contents of the two common plasmid types: *tra* genes were over-represented in F29:A-:B10 plasmids especially *traV*, *traJ*, *traM and traY*. These genes mostly function in conjugal transfers, pilus assembly and production (Skyberg *et al*., 2006; Di Laurenzio *et al*., 1992). Plasmid segregation protein, *parM* was observed only in F29:A-:B10 plasmids while partition gene A, -*parA* was common in F51:A-:B10 plasmids. These genes promote the effective separation of newly replicated cells to two daughter cells during cell division. Their roles hinder post segregational killing of the cells, thereby enhancing plasmid stability. Virulence regulatory genes, *virF* and *virB,* responsible for the control of plasmid encoded virulence genes were prevalent only in F51:A-:B10 plasmids. Numerous toxin-antitoxin (TA) genes were common in the two plasmid types, but the TA genes were over-represented in F29:A-:B10 plasmids. The many toxin-antitoxin (TA) systems affected the stability of the plasmids especially F29:A-:B10 plasmids that was noted to encode higher number of TA systems. However, the cells were able to survive due to other

encoded TA systems such as ccdA and ccdB that ensure that even though the cell do lose its plasmid, it does not experience the effects of the toxin it encodes. It does this by neutralizing the ccdB toxin, thus, effectively blocks the toxin activation when the plasmid is lost. (Ramisetty, and Santhosh, 2016; Bernard and Couturier, 1992; Ogura and Hiraga, 1983). However, the presence of these genes still promoted the plasmid propagation and maintenance (chapter 3).

5.7 Conclusion

Comprehensive insights into the genomic diversity of plasmids in the three phylogroup B2 lineages, because of the differences in their antimicrobial profiles, required analysing them through computational tools and laboratory investigations. These ensure that crucial questions on their adaptation, evolution, pathogenesis, and survival in bacterial host population are adequately addressed. In this study, we examined *E. coli* isolates obtained from human blood, faeces, and urine. Whole genome sequencing was on 335 isolates. These data were used to determine the type and number of plasmids, serotypes, *fimH* types, Incompatibility groups, F-plasmid sequence types, conjugation systems, integron structures, antimicrobial resistance genes and virulence factors associated with the three lineages. We then determined the rate at which the plasmids within these genomes are being transferred, how long the plasmids can persist within the *E. coli* host (plasmid stability), effects of the plasmids on the host (host fitness), and finally, the evolutionary history of the two common plasmid types. This research shows distinct genetic diversity in the plasmids of the phylogroup B2 lineages. Despite that most ST73 strains are less likely to carry conjugative plasmids, strains with F51:A-:B10 plasmids encode many integrons while ST131 strains with

F51:A-:B10 plasmids have integrons that encode greater number of antibiotic resistance determinants. In addition, host cell genetic background has no effect on rate of plasmid transfer, but small differences observed in relation to F-plasmid sequence type can be explained by the presence of multiple *tra* genes in the F-plasmid sequence types. However, host genetic background and plasmid encoded F-plasmid sequence type influenced the plasmid stability and fitness cost associated with plasmid carriage. Analysis of the two common F-plasmid sequence types in the plasmids of the three ST lineages revealed distinct differences in their plasmid characteristics and gene contents. These diversities can be explained by the differences in encoded genetic elements and host genetic background. Despite the differences, the encoded genes play multifunctional roles in the plasmid maintenance and host cell adaptation and evolution.

The success of plasmid-bacterium associations in ST73 and ST95 that made them globally pathogenic even though antibiotic susceptible can be attributed majorly on the genetic content of the plasmids they encode, specifically, the F-plasmid sequence types, F51:A-:B10 and F29:A-:B10. The genes encoded in these plasmid types are majorly responsible for the plasmid propagation, maintenance and possibly host fitness. These genes are highly pathogenic, stable and exert minimal effects on the bacterial host that harbour them. However, some strains of these two lineages are antibiotic susceptible because they either lack conjugative plasmids required for self-transfer during conjugation or/and are devoid of necessary mobile genetic elements needed to counteract the effect of antibiotics. This study has therefore demonstrated that plasmids associated with specific F-plasmid sequence type displayed genotypic characteristics indicative of adaptation, pathogenesis, and persistence in ExPEC

diseases. The distinct functions of the *tra* genes encoded by these plasmids in conjugal transfer is pivotal in the adaptation, evolution, and survival of *E. coli* lineages and finally, the effects of F-plasmid sequence type on plasmid stability and host fitness play important roles on the ability of the plasmids to associate with new bacterial hosts and consequently on the evolution of plasmid-mediated antibiotic resistance.

5.8 Limitations and future directions

5.8.1 Fluoroquinolone resistant strains and STs 73 and 95

Some of the strains of ST131 are fluoroquinolone (FQ) resistant, the reasons why this resistance is so uncommon in ST95 and ST73 is not known In addition, there is a need to investigate if the FQ resistance impose a great fitness cost in ST95 and ST73. Going forward, FQ resistant isolates can further be selected to test the fitness of the mutant against the parent strain.

5.8.2 In vitro analysis

Another limitation of the study is that the plasmid stability and host fitness experiments were only carried out *in vitro*, using conditions very different from the natural environments of the bacterial hosts. We are recommending several experimental models such as animal models to be used to measure stability and fitness cost under conditions that more closely represent the natural habitat of the host bacteria.

5.8.3 Bacteriophage infection

This study did not consider the activities of phages in conjugal transfer. During bacterial conjugation, a bacteriophage can infect and replicates within the bacterium. This is

often possible because some phages use the conjugative Type 4 secretion system (T_4 SS) as an attachment site for invading the bacterial cell. T4SS-specific phages can select against plasmid-carrying bacteria and have been suggested as a way to prevent plasmid-mediated spread of antibiotic resistance. Going forward, bacteriophage interaction in plasmid transfers should be included in future studies.

5.9 References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Molecular Biology of the Cell. New York: *Garland Science*. p. G: **35**. ISBN 978-0-8153-4072-0.

Bahassi, E. M., O'Dea, M. H., Allali, N., Messens, J., Gellert, M. and Couturier, M. (1999)."Interactions of CcdB with DNA gyrase. Inactivation of Gyra, poisoning of the gyrase-DNA complex, and the antidote action of CcdA". *J Biol Chemistr* **274**, 10936–44.

Baltrus, D. A. (2013). Exploring the Costs of Horizontal Gene Transfer. *Trends Ecol Evol* **28**, 489–495.

Bernard, P. and Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes." *J Mol Biol* **226**, 735-745.

Chen, L., Hu, H., Chavda, K. D., Zhao, S., Liu, R., Liang, H., Zhang, W.,Wang, X., Jacobs, M. R., Bonomo, R. A., Kreiswirth, B. N. (2014). Complete Sequence of a KPC-Producing IncN Multidrug-Resistant Plasmid from an Epidemic *Escherichia coli* Sequence Type 131 Strain in China. *Antimicrob Agents Chemother* **58***,* 2422–2425.

Coque, T. M., Novais, A., Carattoli, A., Poirel, L., Pitout, J., Peixe, L., Baquero, F. and Canton, R., Nordmann, P. (2008). Dissemination of Clonally Related *Escherichia coli* Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15. *Emerg Infect Dis* **14**, 195–200.

Dias, R. C. S., Marangoni, D. V. Smith, S. P. Alves, E. M. Pellegrino, F. L. P. C. Riley, L. W. and Moreira, B. M. (2009). Clonal Composition of *Escherichia coli* Causing Community-acquired Urinary Tract Infections in the State of Rio De Janeiro, *Brazil. Microb Drug Resist* **15,** 303-308*. Microbiol Spectr* **2**, 04.

Di Laurenzio, L., Frost, L. S. and Paranchych, W. (1992). "The *TraM* protein of the conjugative plasmid F binds to the origin of transfer of the F and ColE1 plasmids. *Mol Microbiol* **6**, 2951- 2959.

Doyle, M., Fookes, M., Ivens, A., Mangan, M. W., Wain, J. and Dorman, C. J. (2007). An H-NS-like stealth protein aids horizontal DNA transmission in Bacteria. *Science* **315**, 251–252.

Fratamico, P. M., DebRoy, C., Liu, Y., Needleman, D. S., Baranzoni, G. M. and Feng, P. (2016). Advances in Molecular Serotyping and Subtyping of *Escherichia coli*. *Front Microbiol* **3**, 7-644.

Frost, L. S., Ippen-Ihler, K. and Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**, 162–210.

Forde, B.M., Roberts, L.W., Phan, M. D. Peters, K. M*.,* Fleming, B. A*.,* Russell, C. W*.,* Lenherr, S. M., Myers, J. B., Barker, A. P., Fisher, M. A., Chong, T. M., Yin, W. F*.,* Chan, K. G., Schembri, M. A., Mulvey, M. A. and Beatson, S. *A.* (2019) Population dynamics of an *Escherichia coli* ST131 lineage during recurrent urinary tract infection. *Nat Commun* **10,** 3643.

Golding, G. R., Persaud, N., Levett, P. N., McDonald, R. R., Irvine, J., Nsungu, M., Woods, S., Khan, M., Mataseje, L. F. and Mulvey, M. R. (2012). Characterization of *Escherichia coli* Urinary Tract Infection Isolates in Remote Northern Saskatchewan Communities: the Northern Antibiotic Resistance Partnership. *Diagn Microbiol Infect Dis* **74**, 242–247.

Gordon, D. M. (2010). Strain typing and the ecological structure of *Escherichia coli*. *J. AOAC International* **93**, 974–984.

Gordon, D. M., Geyik, S., Clermont, O., O'Brien, C. L., Huang, S., Abayasekara, C., Rajesh, A., Kennedy, K., Collignon, P., Pavli, P., Rodriguez, C., Johnston, B. D., Johnson, J. R., Decousser, J. W. and Denamur, E. (2017). Fine-Scale Structure Analysis Shows Epidemic Patterns of Clonal Complex 95, a Cosmopolitan Escherichia coli Lineage Responsible for Extraintestinal Infection. *mSphere*. pp **2**.

Jang, J., Hur, H. G., Sadowsky, M., Byappanahalli, M., Yan, T. and Ishii, S. (2017). Environmental *Escherichia coli*: Ecology and Public Health Implications-A Review*. J Appl Microbiol*, **123**, 570-581.

Johnson, T. J., Danzeisen, J. L., Youmans, B., Case, K., Llop, K., Munoz-Aguayo, J., Flores-Figueroa, C., Aziz, M., Stoesser, N., Sokurenko, E., Price, L. B. and Johnson, J. R. (2016). Separate F-Type Plasmids Have Shaped the Evolution of the *H*30 Subclone of *Escherichia coli* Sequence Type 131. *Clinical Science and Epidemiology mSphere,* **1** (4).

Johnston, C., Martin, B., Fichant, G., Polard, P. and Claverys, J. P. (2014). "Bacterial Transformation: Distribution, Shared Mechanisms and Divergent Control". *Nature Reviews Microbiology* **12**, 181–96.

Kallonen, T., Brodrick, H. J, Harris, S. R, Coriander, 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**, 437-1449.

Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004). Pathogenic *Escherichia coli. Nat Rev Microbiol* **2**, 123–140.

Kraft, F. and Kurth, I. (2020). Long-read sequencing to understand genome biology and cell Function. *The Inter Jour Biochem and Cell Biol* **126**, 105799.

Lan, R., Lumb, B., Ryan, D. and Reeves, P. (2001). Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli. Infect Immun* **69**, 6303-6309.

Lawley, T. D., Klimke, W. A., Gubbins, M. J. and Frost, L. S. (2003). F- Factor Conjugation is a True Type IV Secretion System. *FEMS Microbiol Lett* **224**, 1–15.

Lee, S. W. and Edlin, G. (1985). Expression of tetracycline resistance in pBR322 derivatives reduces the reproductive fitness of plasmid-containing *Escherichia coli*. *Gene* **39**, 173-180.

Liu, Y. Y., Wang, Y., Walsh, T. R., Yi, L. X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L. F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J. H. and Shen, J. (2016). Emergence of Plasmid-Mediated Colistin Resistance Mechanism MCR-1 in Animals and Human Beings in China: a Microbiological and Molecular Biological Study. *Lancet Infect Dis* **16**, 161-168.

Nicolas-Chanoine, M. H., Bertrand, X. and Madec, J. Y. (2014). *Escherichia coli* ST131, an Intriguing Clonal Group. *Clin Microbiol Rev* **27**, 543–574.

Nicolas-Chanoine, M. H., Blanco, J., Leflon-Guibout, V., Demarty, R., Alonso, M. P., Caniça, M. M., Park, Y. Y., Lavigne, J. P., Pitout, J. and Johnson, J. R. (2008). Intercontinental Emergence of *Escherichia coli* Clone O25:H4-ST131 Producing CTX-M-15. *J Antimicrob Chemother* **61**, 273-281.

Ogura, T. and Hiraga, S. (1983). "Mini-F plasmid genes that couple host cell division to plasmid proliferation." *Proc Natl Acad Sci* U.S.A. **80**, 4784-4788.

Quainoo, S., Coolen, J. P.M., van Hijum, S. A. F. T, Huynen, M. A., Melchers, W. J. G., van Schaik, W. and Wertheim, H. F. L. (2017). Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev* **30**, 1015-1063.

Ramisetty, B. C. and Santhosh, R. S. (2016). "Horizontal gene transfer of chromosomal Type II toxin-antitoxin systems of *Escherichia coli*". *FEMS Microbiology Letters* **363**, 238.

Saavedra De Bast, M., Mine, N. and Van Melderen, L. (2008). "Chromosomal toxin-antitoxin systems may act as antiaddiction modules". *J Bacteriol* **190,** 4603–9.

Shachrai, I., Zaslaver, A., Alon, U. and Dekel, E. (2010). Cost of Unneeded Proteins in *E. coli* is Reduced after Several Generations in Exponential Growth. *Mol Cell* **38**, 758–767.

Sherley, M., Gordon, D. M. and Collignon, P. J. (2003). Species Differences in Plasmid Carriage in the Enterobacteriaceae. Short Communication. *ScienceDirect* **49**, 79-85.

Skov, R. L. and Monnet, D. L. (2016). Plasmid-Mediated Colistin Resistance (mcr-1 gene): Three Months Later, the Story Unfolds. *Euro Surveill* **21**, 30155.

Skyberg, J. A., Johnson, T. J., Johnson, J. R., Clabots, C., Logue, C. M. and Nolan, L. K. (2006). "Acquisition of avian pathogenic Escherichia coli plasmids by a commensal E. coli isolate enhances its abilities to kill chicken embryos, grow in human urine and colonize the murine kidney." *Infect Immun* **74**, 6287-6292.

Stearns, S. C., and Hoekstra, R. F. (2005). Evolution: An introduction (2nd ed.). Oxford, NY: Oxford Univ. Press. pp. 38-40.

Stephens, C. M., Adams-Sapper, S. Manraj Sekhon, Johnson, J. R. and Riley, L. W. (2017). Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. *mSphere*, **2** (2). DOI: 10.1128/mSphere.00390-16.

Warnes, A. and Stephenson, J. R. (1986). The insertion of large pieces of foreign genetic material reduces the stability of bacterial plasmids. *Plasmid* **16**, 116-123.

World Health Organization (2020). Global antimicrobial resistance and use surveillance System (GLASS) Report - Early Implementation 2020. Geneva, Switzerland.